LUND UNIVERSITY

A Naturalistic Approach to the Generality Problem

Olsson, Erik J

Published in:
Goldman and His Critics

DOI:
10.1002/9781118609378.ch8

2016

Link to publication

Citation for published version (APA):
Olsson, E. J. (2016). A Naturalistic Approach to the Generality Problem. In B. P. McLaughlin, & H. Kornblith (Eds.), *Goldman and His Critics* (pp. 178-199). (Philosophers and Their Critics; Vol. 16). Wiley-Blackwell. https://doi.org/10.1002/9781118609378.ch8

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Download date: 06. Oct. 2026

A Naturalistic Approach to the Generality Problem

To appear in H. Kornblith and B. McLaughlin (eds.), *Alvin Goldman and His Critics*, John Wiley and Sons, Inc.: Hoboken, NJ.

Erik J. Olsson
Lund University

1. Introduction

Process reliabilism, or reliabilism for short, is the view that S knows that p if and only if (i) p is true, (ii) S believes that p, (iii) S's belief that p was acquired through a reliable process, and (iv) an appropriate anti-Gettier condition is satisfied. Reliabilism is sometimes advocated, alternatively, as a theory of epistemic justification, the main idea being that a person is justified in belief that p just in case her belief that p was formed via a reliable process.[1] For the purposes of the following discussion, there is no need to make a sharp distinction between these two brands of reliabilism. For definiteness, I will, initially, state much of the discussion in terms of knowledge.

The source of the generality problem for reliabilism is the observation that, because a process token is an unrepeatable causal sequence occurring at a particular time and place, it makes no good sense to ask whether a token process is reliable in itself. Rather, what can fundamentally be reliable or not are process *types*. For instance, the concrete process of Jones's coming to believe that he won the lottery on May 1, 2007, is itself neither reliable nor unreliable. However, given that its associated type is taken to be "belief formed through reading the local newspaper", it is (probably) reliable. A process token can still be said to be (un)reliable in a derivative sense if its associated process type is (un)reliable.

[1] The process reliabilist account of knowledge was originally formulated by Ramsey (1931). See Olsson (2004) for a discussion. For a modern (post-Gettier) account, see Goldman (1986). The process reliabilist theory of justification was first put forward in Goldman (1979).

The problem now arises, more specifically, because each token process can be classified as belonging to a great many different types, and it is not obvious how to single out one of these types rather than another as *the* unique associated type of the process in question. For example, the process leading up to Jones's belief could be classified narrowly as belonging to the type whose sole member is Jones's coming to have his belief about the lottery, or, to take the other extreme, broadly as a belief formed through reading.

Furthermore, depending on what type is singled out as special we may get different verdicts as regards the reliability of the process in question. Given the narrow classification in terms of the type whose sole member is the process producing Jones's belief, that process will, if the belief is true, be reliable. If it is seen instead as instantiating the general type "reading", it might be judged unreliable. Reading in general, irrespective of what is being read, is probably not reliable to an extent that suffices for knowledge.

These considerations reveal what appears to be an unacceptable lacuna in the reliabilist account of knowledge. From the reliabilist's perspective, whether a person knows or not will in many cases depend on whether the type of process producing the belief in question is reliable. And yet, as most commentators would agree, reliabilists have generally failed to clearly identify *the* type pertaining to a given token. In the absence of a principled account for how to select the relevant type, the reliabilist theory appears to be, in Conee and Feldman's words, "radically incomplete" (1998, p. 3).[2]

Feldman (1985) and Conee and Feldman (1998) cite a number of criteria they think any acceptable solution should satisfy, beyond associating every process token with a corresponding process type.[3] According to Feldman, solving the

[2] The generality problem is of course not the only problem facing reliabilism. There is also, for instance, the value problem: the problem of explaining how reliabilist knowledge can be more valuable than mere true belief. I have had the privilege of writing a joint paper with Goldman on that topic (Goldman and Olsson, 2009), in which we propose several alternative, though compatible and perhaps ultimately complementary, solutions. I have developed my own preferred approach in several other publications. See Olsson (2007, 2008, 2009, 2011), and Olsson and Jönsson (2011).

[3] The generality problem is usually stated as one of finding a *unique* relevant process type for each process token. Logically speaking, however, it would suffice to identify, for each token, a *class* of

generality problem for reliabilism requires showing how to avoid what he calls the single case and no distinction problems. The single case problem occurs when a process type is described so narrowly that only one instance of it ever occurs, and hence the type is either completely reliable or completely unreliable depending on whether the belief is true or false. The no distinction problem arises when beliefs of obviously different epistemic status are produced by tokens that are of the same (broad) relevant type. These two limiting cases were illustrated above in our example with Jones's belief about the lottery.

Conee and Feldman lay down three additional requirements for a solution to the generality problem. First, a solution must be "principled" in the sense of not being made on an *ad hoc* basis. Second, the rule must make reasonable epistemic classifications, by which is meant that the types identified must have a reliability that is plausibly correlated with the justificational status of the beliefs in question. Third, a solution must remain true to the spirit of the reliabilist approach and not characterize the relevant type of process in epistemic terms that are alien to reliabilist theorizing.[4]

Reliabilists have not been insensitive to this problematic which was identified, for example, by Alvin I. Goldman in his classic 1979 paper and is portrayed as a serious issue for reliabilism in Goldman (1986) and, recently, in Goldman (2008). It is now considered to be a main challenge for a reliabilist theory by reliabilists and their critics alike.[5] Goldman has given various cues for how he would like to address this problem but it is my understanding that he has not yet converged on a specific solution, and in Goldman (2008) he expresses dissatisfaction with the approach defended in his earlier work:

---

types whose members are either all reliable or all unreliable. Nevertheless, I will follow the mainstream and assume that the problem is to find a unique type for each token.

[4] A proposal referring to natural kinds is made in Alston (1995). Mark Heller (1995) offers a contextualist approach. For other approaches to the generality problem, see Brandom (1998), Adler and Levin (2002), Wunderlich (2003), Comesaña (2006) and Kappel (2006). See also the exchange between Lepin (2007) and Christensen (2007).

[5] The generality objection can be found in many surveys of contemporary epistemology, e.g. Hetherington (1996), pp. 40-41, Lycan (1998), pp. 110-111, Plantinga (1993), p. 198, Pollock (1986), pp. 118-120, and Pollock and Cruz (1999), pp. 116-118. For a recent example, see Lemos (2007), pp. 92-94.

> Which repeatable type should be selected for purposes of assigning a determinate reliability number to the process token? "What Is Justified Belief?" does not resolve this question, and it remains an important one. Goldman (1979) says that cognitive processes should be restricted in "extent" to events within the organism's nervous system (although he does not abide by this restriction in some of his own illustrations of process types). But this restriction provides no criterion for pinpointing a unique process type. It appears, however, that a determinate reliability number cannot be assigned to a process token unless a unique type is selected.[6]

I have found no reason to disagree with this self-critical assessment of Goldman's. In the same overview of reliabilism, Goldman goes on to discuss several proposals by other authors in an open-minded spirit, e.g. that advanced by Beebe (2004).

In this paper I will identify, in section 2, what is in my view the main challenge for reliabilism in relation to the typing of belief forming processes. This challenge was raised by Conee and Feldman in their 1998 article. I will then proceed, in section 3, to undermine that challenge by drawing on insights in cognitive science in a way that should make this response attractive to practitioners of naturalized epistemology, a group to which Goldman famously belongs.[7] Nevertheless, Goldman has, in our private correspondence, offered a number of interesting objections to what I am about to propose. Part of the purpose of the present paper is to address what I take to be his main worries in that regard, including what he terms his "principal objection" (personal communication).

2. Conee and Feldman's main challenge

Suppose that the critic is correct in thinking (A) that there are generally many different ways of classifying a belief formation process, and (B) that depending on

[6] Cf. Goldman (1986), pp. 49-51.
[7] The approach advocated here was first proposed in Olsson (in press).

how the process is classified we will get different verdicts as to whether or not the process was reliable. Let us finally also grant (C) that the reliabilist has so far been unable to devise an explicit rule by means of which the right type can be identified in a given case. In what sense would these be disturbing facts for the reliabilist? Conee and Feldman's answer, of course, is that these facts together turn reliabilism into a "radically incomplete" epistemological theory. But what, exactly, is the reasoning leading up to this conclusion? Extracting a clear answer to this question from Conee and Feldman's work is surprisingly difficult.

Perhaps Conee and Feldman are taking reliabilism to provide not merely a conceptual analysis of knowledge or justification but also a normative method for deciding, in practice, whether someone knows or is justified. There are some indications in this direction in their 1998 paper, in which they write, for instance, that "[o]nly when a bearer of reliability has been identified does the theory have any implications about the justification of beliefs in particular cases" (p. 3). From this point of view, their criticism is relatively easy to appreciate: reliabilism is radically incomplete because it fails to give a method for identifying a process type "in particular cases". However, this interpretation of Conee and Feldman renders their criticism largely uninteresting for the simple reason that reliabilism is usually not advanced as a decision method along these lines, and in his 1985 paper Feldman himself explicitly distances himself from this understanding of reliabilism and the generality problem, writing "I do not assume that an acceptable solution to The Problem of Generality must provide a practical and useful method for identifying relevant types" (p. 173, footnote 6).

There is another, perhaps deeper, reason why this rendering of the generality problem is dubious. Critics of reliabilism universally present the generality problem as a challenge exclusively facing that particular theory. It is supposed to tell against reliabilism in a way that should make us more favorably disposed toward its main competitors, e.g., variants of JTB, virtue theory or, in Conee and Feldman's case, evidentialism. This can be so only if there is no corresponding problem arising for those other well-established analyses of knowledge. But if the complaint is that reliabilism does not provide a method for deciding in practice

whether someone knows, then it is doubtful whether its critics have indeed succeeded in identifying a problem unique to reliabilism. After all, JTB analyses of knowledge, to take one example, do not provide such a method either because they typically do not provide a method for deciding, in particular cases, whether someone is justified in her belief. Comesaña (2006), similarly, identifies a problem for evidentialism analogous to the generality problem for reliabilism. Providing, in Feldman's words, practical and useful methods of that kind is an interesting task in its own right but it is one that is orthogonal to the problem of providing a conceptual analysis of knowledge or justification and – returning to the main point – one that concerns just about any account of those concepts that comes to mind.[8]

The upshot is that to the extent that Conee and Feldman's attack on reliabilism rests on a failure to appreciate the two points just made, it can be dismissed by reliabilist as not seriously threatening their position.

However, there are other elements of Conee and Feldman's critique that cannot be so readily discarded. Their paper also contains an ingenious argument that directly targets the reliabilist theory, in so far as its practitioners attempt to capture, at least approximately, the folk concepts of knowledge and justification (and, to be fair, most of us do). The most stimulating part of Conee and Feldman's attack can charitably be viewed as targeting the notion that the reliabilist concepts of knowledge and justification are sufficiently similar to their folk counterparts.[9]

[8] The generality problem may also be thought of as the normative problem of how to resolve conflict concerning the proper categorizations of belief forming processes. Suppose X types process token t in one way whereas Y types that same token in another way, and that the effect is that X and Y differ in their reliability assessment. The problem is whether there is a rational way for X and Y to come to an agreement on how to type the process, e.g. by adopting a common type representing "suspension of judgment". Intriguing as this problem is, it arises in principle for any epistemological theory, and not just for reliabilism. With regard to any theory of justification, for instance, we may ask how it proposes to resolve conflicting judgments regarding justification.

[9] Carnap should be credited for an unusually clear and plausible formulation of the sense in which a conceptual analysis should be faithful to ordinary usage. According to one of his requirement on a good philosophical "explication", "[t]he explicatum [the thing which explicates] is to be *similar to the explicandum* [the thing to be explicated] in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used; however, close similarity is not required, and considerable differences are permitted" (Carnap 1950, p. 70).

Let us focus for simplicity on the reliabilist theory of justification (RJ) and on Conee and Feldman's criticism as targeting the faithfulness of RJ with regard to our everyday concept of justification. RJ can fail in this regard in two main ways: in being either too broad or too narrow. RJ is too broad if beliefs can be produced by reliable processes without being justified. It is too narrow if beliefs can be justified without being produced by reliable processes.

As for the latter case, suppose there are clear cases of justification, i.e. cases where we would all agree that a subject S is justified in believing that p. Imagine, for instance, that we are presented with a picture of Wilma standing in front of a tree in broad daylight. Wilma is, for all we can see, looking at the tree. There is nothing obscuring here view, we have no reason to think she is blind, and so on. Clearly we would then all assent to Wilma's being justified in believing that there is a tree in front of her. This is a clear case of justification. But is it also a clear case of belief acquisition through a reliable process? Given claims (A) – (C) above one might be led to think that it is not: given that there are many different ways to classify Wilma's belief formation process, leading possibly to widely different reliability assessments, and given that we have no explicit rule that we can appeal to in the selection of one classification rather than another, it would seem that Wilma's is not a clear case of reliable belief formation. Rather, it will be, in a sense, indeterminate whether or not Wilma's belief was reliably formed.

This is indeed what certain theoretical considerations in the form of (A) – (C) suggest. But why should we rest content with theory when we can study what happens in practice? Here is a sketch of a proposal for how the matter could be tested experimentally:

The proposal involves two groups of subjects confronted with, say, twenty episodes of ordinary life involving a person coming to believe something for some reason or other (like Wilma above). For instance, the episodes could be presented to the subjects as film sequences.[10] Each subject in one of the groups is asked to state independently for each episode shown whether the person figuring in that episode is justified in her belief. Let us refer to this group as the *justification*

[10] I am indebted to Martin L. Jönsson for the proposal to use film sequences for this purpose.

group, or the J-group for short. Each subject in the other group is asked to state independently for each episode shown whether the person figuring in that episode acquired her belief in a reliable way. Let us refer to this group as the reliability group, or R-group for short. Obviously, the instructions for the subjects participating in this experiment should not contain any linguistic hints that may make them favor one classification rather than another.[11]

Assume now that this experiment has been carried out and that we got a good positive match between the reports of the J-group and the reports of the R-group. In other words, more or less the same the same episodes are described, by all or almost all members of the J-group, as involving justification and by all or almost all members of the R-group as involving reliable belief acquisition. That would be clear evidence in favor of the reliabilist theory of justification with regard to its similarity to our everyday concept of justification.

What outcome would *disconfirm* that aspect of the reliabilist theory of justification? If we were to carry out the same experiments only to find there to be little or no positive correlation between judgments of justification and judgments of reliability that would be an unfortunate fact for the reliabilist theory, as it would suggest that the reliabilist concept of justification lacks sufficient similarity to the folk concept of justification.[12]

For the purposes of the following discussion we need to distinguish between two kinds of disconfirmation of RJ. We would have a clear case of disconfirmation if, for instance, whenever all or almost all members of the J-group agreed that the person in the episode is justified, all or almost all members of the R-group agreed that the person did not acquire her belief in a reliable fashion. Let us refer to this as *category I disconfirmation*. But we could also encounter an

[11] We would like to use two groups in order to avoid various framing problems. If one group were to assess both justification and reliability, say in that order, the members' assessment of justification might conceivably influence their assessment of reliability.

[12] Similarly, a reliabilist theory of *knowledge* would be confirmed, with regard to faithfulness to ordinary language, if, in a significant number of cases, whenever all or almost all members of one group of subjects agrees that the person in the episode knows, all or almost all members of the reliability group agrees that the belief in question was reliably acquired. By contrast, if, in a significant number of cases, the members in the knowledge group agreed that the person knows, without the members of the second group reporting that the person's belief was reliably acquired, that would disconfirm the reliabilist theory of knowledge.

outcome in which all or almost all members of the J-group agree in their justification assessments but in which there is strong dissonance among the members of the R-group concerning reliability. For instance, some members of the R-group may insist that the process was reliable, others may be equally convinced that it was unreliable and still others may fail to reach a verdict at all. Let us refer to this as *category II disconfirmation*.

Let us ponder on the likelihood of various alternative scenarios. As I construe them, Conee and Feldman report some rather strong views on this matter, to the effect that whereas the subjects in the J-group will, at least in paradigmatic cases, tend to give the same reports, the subjects in the R-group are likely to give widely unsystematic responses. In other words, Conee and Feldman predict that the experiment will result in category II disconfirmation of RJ.

Why do Conee and Feldman predict that the subjects in the J-group will tend to give the same reports in paradigmatic cases? In our experiment, the episodes displayed to the participating subjects may be selected so as to exemplify what Conee and Feldman call "typical contexts in which knowledge and justification claims have clear truth value" (1997, p. 24). Conee and Feldman are committed to there being such contexts, and we must interpret them to be assuming that people will give more or less identical reports when asked whether knowledge or justification is present in those contexts.

Conee and Feldman's argument for thinking that the subjects in the R-group will fail to give converging reports is more subtle. Here is a step-by-step reconstruction:

(CF1) People sometimes agree on typing and reliability but this happens only if the given type they converge upon is salient in the conversational context.

(CF2) The only way in which a type may be become salient in a conversational context is by means of linguistic presentation.

(CF3) Ordinarily, no type is linguistically presented.

(CF4) Hence, ordinarily, no type is salient and so people will disagree on typing and reliability.[13]

This train of thought applies to our experiment because we were careful to assume that no type is linguistically indicated in the instructions to the participating subjects. Hence, Conee and Feldman offer the predication that the subjects in the R-group will be unable to converge on the same reliability assessment in concrete cases and that this failure is due to the lack of salient types. For instance, the number of R-group subjects reporting Wilma's belief formation process to be reliable may be roughly the same as the number of subjects reporting it to be unreliable.

My own prediction differs significantly from Conee and Feldman's. I anticipate that both groups will report in a homogeneous manner, and there will be significant positive correlation between the reports of the two groups. For instance, most subjects in both groups will think of Wilma as simply 'seeing a tree', and this will make most subjects in the J-group report that Wilma is justified in her belief about the tree, and most subjects in the R-group report that the process by means of which Wilma formed her belief is reliable. I will proceed now to add substance to these claims.

[13] See Conee and Feldman (1998), p. 22-23, where they argue against Heller's claim that the relevant process type will be contextually determined in many situations in which knowledge or justification is attributed: "There are some situations in which phrases referring to process tokens apparently work in the way Heller describes. For example, suppose Jones says, 'I have three ways to start my old jalopy: first, shifting into gear while rolling it down a hill; second, jump-starting it; and third, praying and then turning the key. Only the first two usually work.' Suppose that Jones then starts his car by jump-starting it. He remarks:
P. 'The process by which I just started my car is reliable'.
Here, Jones's explicit mention of the three types serves to limit drastically the types under consideration. The token mentioned in (P) is of one of those types only … In typical knowledge attributions, however, no contextual narrowing of candidate process types occurs … Ordinarily, no class of types of belief-forming processes will have been made contextually salient. And nothing else about typical contexts isolates any one type." See also Feldman and Conee (2002), p. 102, footnote 1.

3. Conee and Feldman's challenged undermined

One reason why we should expect people to give the same reliability verdicts in concrete cases comes from our ordinary life experience, for we often agree, it seems, about what is reliable and not, even in the complete absence of linguistic classificatory pointers. If someone comes forward claiming that this car is reliable, pointing to a new Mercedes Benz, you would presumably tend to agree that it is. Similarly, we agree that the process by means of which the heart pumps blood through our veins is reliable (if, alas, only up to a point). Or suppose that we discover Karen sitting in the university library reading the Encyclopedia Britannica. Karen now raises her head saying, "The way I just learned the molecular structure of gold is reliable". Few of us would be inclined to object.

There is another reason for expecting that people should tend to converge on matters of reliability even if no relevant type-classification is made salient by linguistic means, for this is what influential empirically-based work in cognitive psychology strongly suggests. In underpinning this claim I will draw on work on salience and so-called basic level effects in the literature on categorization.

Several psychological studies have been conducted on how people classify events and, in particular, what it is that makes some event categorizes more natural or salient. This work is relevant here because belief forming processes are kinds of events. Zacks and Tversky (2001) is a useful overview of the following and other advances in the psychological literature on event categorization. One line of research in this area indicates that what event type becomes salient may depend on the *time scale*. Consider "crossing the street", "walking to school", and "getting an education" as three different types corresponding to the same token event. According to the findings, event types corresponding to a medium range time scale, such as "walking to school" are more likely to be salient than other more extreme types. This effect does not depend in any way on one or more types being made salient by linguistic means.

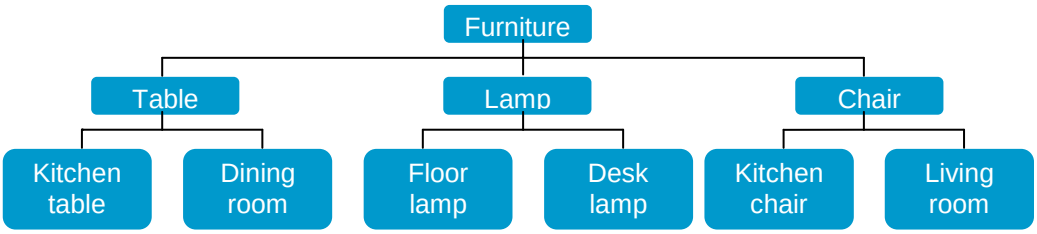
Here is an even more suggestive piece of evidence: According to one influential theory recognizing an event as an instance of a category consists of (A)

matching it to a *schema* stored in memory and (B) matching features of the world to variables in the schema. Example of a schemata might be "X walks to Y", "X robs Y", and so on. Some schemata may be of evolutionary origin, others are developed in a social context through a learning process. The crucial point however is that an event type may become salient because it matches and activates a schema stored in memory. This phenomenon, too, occurs even if no type is singled out linguistically.

This is not the place to go into the details of how these phenomena transpire. What can be said with some confidence is that there is a general tendency in the relevant part of cognitive psychology to regard with suspicion any theory – such as that advocated by Conee and Feldman – according to which a type can become salient in a given context only through being explicitly mentioned.

So far we have seen some evidence for thinking that non-linguistic factors can have salience-making powers. Such evidence does not by itself indicate that we should expect people to converge in their categorization of events of belief acquisition, for presumably most theories of salience, including those highlighting the role of non-linguistic mechanisms, will involve criteria that are too liberal to pick out unique categories. Suppose, for example, that we have roughly the same structured process schemata stored in memory and that they serve to narrow down the psychologically real categories, in any given context, to just a few especially prominent ones. Little would prevent a given token process from potentially activating more than one such schema. For instance, Smith's seeing a tree could conceivably activate the following schemata all of which have some claim to psychological reality: *X perceives Y*, *X sees Y*, *X sees Y in bright sunlight*. What is it, then, that makes us zoom in, eventually, on, say, *X sees Y*?

The forgoing remarks raise the question as to how people can come to agree on a natural *classificatory level* in taxonomies of salient types. The starting point from which will gradually approach an answer to that question will be object, rather than event, categorization. Consider the taxonomy in Figure 1.



Figures 1: Taxonomy for furniture

Experimental studies of taxonomies as the one in Figure 1 have established that there is a level of categorization that is "basic" in the sense of being most natural. Thus, people will tend to think of things in terms of middle range categories like *Table*, *Lamp* and *Chairs* rather than in terms of superordinate (*Furniture*) or subordinate (*Kitchen chair* etc.) categories. Evidence for a basic middle level comes from a variety of experimental sources: basic level categories tend to be named when people are shown an object; in recognition tasks, people recognize basic level objects faster than either subordinates or superordinates; basic-level names generally have arisen earlier in the development of languages; basic categories are used earlier in the naming and other behavior of young children; finally, basic-level names tend to be shorter and more frequently used than names of superordinate or subordinate categories.[14]

The standard explanation of this so-called basic level effect is based on the observation that categories tend to be associated with "attributes" which are things that can be inferred (in a defeasible sense) from membership in the category. For example, if a thing is classified as a *Chair*, we can infer that we can sit on it. If it classified as a *Lamp*, we can infer that it can be lit, and so on. Rosch et al (1976), perhaps the single most important contribution to this area of research (as judged by the number of citations in ISI Web of Knowledge), lists the following culturally significant attributes of furniture categories:

[14] Cf. Rogers and Patterson (2007), p. 451: "… basic level effects are among the best known and most robust phenomena in the study of human categorization."

Furniture: no attributes

Chair: legs, seat, back, arms, comfortable, four legs, wood, holds people – you can sit on it

Kitchen chair: no additional

Living room chair: large, soft, cushion (additional attributes)

Thus, given that something is a chair, it can be inferred that the thing has legs, a seat, and so on. Given that something is a kitchen chair, the same inferences can be made and no new inferences are added. From the categorization of something simply as a piece of furniture, no useful conclusion can be drawn.

We can now explain the basic level effect by invoking two cognitive mechanisms assumed to be universally operative in human categorization. One mechanism favors economy of representation and the other usefulness (informativity) of categorization. We can conveniently speak of these mechanisms as representing two cognitive goals. The goal of economy requires that the classifier treat as many things as possible as "equivalent", i.e. as belonging to the same category. This goal will tend to favor broad categories. The goal of usefulness requires that the classifier seek a maximally useful classification of a given thing. A classification is useful to the extent that interesting attributes of the object can be inferred (again defeasibly) from category membership alone. This goal will tend to favor narrow categories.[15] The challenge facing a classifying subject is to make a reasonable trade-off between these two goals, i.e., to choose a type that is informative and yet cognitively parsimonious.

Rosch and her associates proposed a specific way of striking the right balance: the basic classificatory level, they wrote, is "the most inclusive level in a taxonomy at which a cluster of attributes, believed to be common to the class named, would be listed" (Rosch et al, 1976, pp. 435-436, original emphasis removed). Let us refer to this rule for identifying the basic level as the *Rosch rule*. For instance, the *Chair* level is the basic level because it involves a cluster of attributes which are not listed at higher levels in the taxonomy. The *Kitchen Chair*

[15] See Rosch et al (1976), pp. 383-385, for details.

level is not basic because, although it too involves a cluster of listed attributes, those attributes are listed at the higher *Chair* level as well.[16] [17]

I will now propose to treat the classification of belief forming processes in an analogous fashion. The proposal is to use the Rosch rule to determine the type for a given process token just as we did for concrete objects. This is in line with the observation of basic level effects in the categorization of events (of which belief formation processes are a special case) and the claim made by Rosch and her colleagues that the principle of category formation is universally valid. The two cognitive goals previously alluded to are operative in the categorization of belief formation processes, so that our natural inclination is to obtain a categorization that is both economical and informative, just as before.[18] The most basic level in a process taxonomy is the most inclusive level at which a cluster of attributes, believed to be common to the class named, would be listed.

The following example illustrates how this is supposed to work.

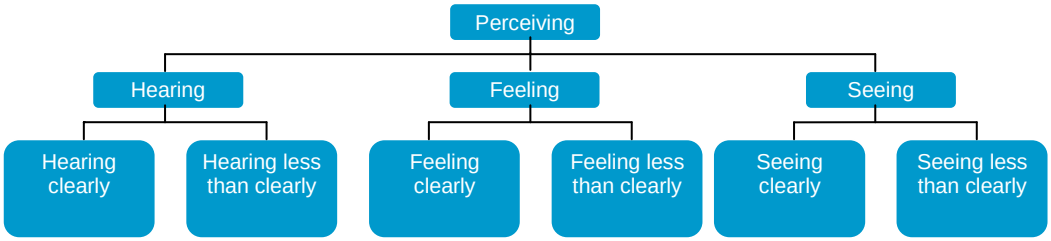


Figure 2: Hypothetical taxonomy for belief formation

[16] The Rosch rule suffers from imprecision concerning how many attributes it takes to form a "cluster". This potential shortcoming turns out to be inconsequential in the present context, as we will focus on a special case of the rule in which the notion of an attribute cluster plays no role.
[17] Variants of the explanatory model originally proposed by Rosch and colleagues have continuously been employed in order to explain phenomena connected with categorization and non-logical reasoning. See Jönsson and Hampton (2006, 2008) for recent examples.
[18] Rosch et al (1976): "the principle of category formation … is claimed to be universal" (p. 435). See also Mervin and Rosch (1981), p. 93. Zacks and Tversky (2001) discuss some work confirming the existence of basic level effects in event categorization. [18] I am not aware of any actual experiments designed to test the basic level theory for the special case of belief formation processes. Nor am I, for that matter, aware of any evidence suggesting that such processes should constitute an exception to the general rule that classification of events conforms to that theory.

I hypothesize that the following attributes would be considered to be of potential importance for most people:[19]

Perceiving: no attributes

Seeing: reliable (for most practical purposes), believing, proximity to object, nothing (opaque) obstructing the view

Seeing clearly: very reliable (for scientific/legal purposes)[20]

Seeing less than clearly: no additional[21]

The basic level is given by the most inclusive level classification of visual perception at which a cluster of attributes are listed, which in this case is the level of *Seeing*.

The account we have given of type convergence also predicts convergence in matters of reliability. For the reasons given, people will tend to categorize belief formation processes in terms of *Seeing*, *Hearing*, and so on. They will then tend to infer the attributes associated with those categories.

Let us return to the previous example of Wilma for a concrete illustration. I claimed that most of the subjects in the R-group would think of her as simple "seeing a tree", a process they will report to be reliable (at least for everyday purposes). The reason is that *Seeing* is very plausibly a basic level event category and *Tree* a basic level object category.[22] Either category allows many inferences to be drawn to potentially useful attributes. Combining these categories into "seeing a tree" is a cognitively attractive move because it results in a description that is compact yet loaded with information content.

Someone might object to our explanation of typing convergence on the following basis: "Fine, maybe this succeeds in showing, at least in outline, how we can agree on a process classification and reliability in some cases, but these cases do not seem to correspond to those that epistemologist take special interest in. Some of the 'attributes' of *Seeing* – like 'proximity to the object' – are not

---

[20] "Seeing" is here and elsewhere taken in its non-veridical sense.

[21] Alternatively, "not very reliable (for scientific/legal purposes)" could be added as an attribute of the category *Seeing less than clearly*.

[22] Rosch et al (1976) found empirical support for the claim that *Tree* is a basic level object category (pp. 390-391).

really epistemically interesting, which raises the question why we should base our classifications on them. In epistemic contexts, what we care about is simply the reliability of the process, and that is all there is to it. What is needed is an argument as to why we tend to agree on classification/reliability when the reliability only is at stake. The Rosch-based theory presented here doesn't seem to do that job." So far we have assumed that we are interested in a *multi-purpose classification* where many attributes are (potentially) important.[23] What the objector points out, correctly, is that the original classificatory problem can also be understood as one of *single purpose classification*. On this interpretation, there is only one attribute/property that we care about: reliability. So what happens in this case? The original Rosch rule for multi-purpose classification, we recall, states that the basic level of classification is the most inclusive level in a taxonomy at which a cluster of attributes, believed to be common to the class named, would be listed. This suggests that, in the single purpose case, the basic level of classification should be the most inclusive level in a taxonomy at which the attribute of interest would be listed.

There is however an immediate problem with this proposal. Suppose the taxonomy consists of the superordinate *Being deceived* and the subordinates *Being deceived intentionally* and *Being deceived unintentionally*, and that the token process in question is a case of, say, intentional deception. Since "reliable" is not an attribute of any category to which the token process belongs, there is no level in the taxonomy at which "reliable" is listed. Hence, no type is singled out by our proposed rule.

The problem is quickly solved by adding that the relevant taxonomic levels are such that either the attribute itself *or its negation* is listed. In other words, *the basic level of classification, in the single purpose case, is the most inclusive level*

[23] Cf. Corter and Gluck (1992), p. 293: "Our basic assumption is that there is functional value for a person to have accurate information about the features of things. For example, an organism searching for food needs to know whether a given plant part is poisonous, nutritive, sweet, tough, and so on. Some features of instances may be useful only indirectly – for example, to generate tests to confirm tentative identifications. However, because a person will experience a variety of need states and goals across time, generally, there is value for the person to have accurate information about all the features of instances."

in a taxonomy at which the attribute or its negation would be listed. This solves the problem: since all the categories to which the token process belongs will be associated with the attribute "unreliable", the rule will single out a unique type namely *Being deceived*. That category is the most inclusive type to which the token belongs and which is such that the attribute of interest, or – as in this case – its negation, is listed.

To take another example, the following attribute list would be relevant if the task is single purpose categorization with respect to reliability (for most practical purposes):

Perceiving: no attributes

Seeing: reliable (for most practical purposes)

Seeing clearly: no additional

Seeing less than clearly: no additional

If, as we assume, the process was a case not only of *Seeing* but also of *Seeing clearly*, both these categories satisfy the condition that "reliable (for most practical purposes)" is listed. However, *Seeing* is the most inclusive category satisfying that condition, which is why it is singled out by the single purpose Rosch rule as the type of the process in question.[24]

4. Typing in non-monotonic taxonomies: a response to Goldman's principal objection

So far we have been looking at taxonomies that are *monotonic* in the sense that subordinate categories "inherit" all attributes from their superordinates.[25] In the furniture case, for instance, the *Chair* attribute "has a seat" was supposed to be inherited by the subordinate categories of *Kitchen Chair* and *Living Room Chair*. Likewise, the reliability attribute of *Seeing* was assumed to be inherited by its

[24] The result of single purpose classification might diverge from the result of multi-purpose classification, which is in this case *Seeing*. Take as the single-purpose attribute that of being very reliable. In Figure 2, "very reliable" appears only at the level of *Seeing clearly*, which is therefore the single purpose basic level category.

[25] This is the term used by Corter and Gluck (1992), p. 296.

subordinate categories. Monotonic taxonomies admit a highly compressed cognitive representation which allows us to store attributes only once without having to repeat them for subordinate categories. For that reason, we should expect the human mind to favor storing information in monotonic taxonomies if that is an option. We have provided an account of process typing that should make us confident that people will often converge on the same type for monotonic taxonomies. But what about the non-monotonic case?

Consider the following example from Goldman (personal communication): Suppose that John's process leading up to his belief that the person over there is Susan instantiates the type *Seeing*. This does not prevent it from also instantiating the type *Occluded seeing*, by which is meant that the target of vision is at least partly occluded. If classified as *Seeing*, the process can be inferred to be reliable (for everyday purposes), whereas *Occluded seeing* will be associated with the attribute "unreliable". The matter doesn't end here: suppose that in the case we are discussing, the occlusion only pertains to the target's periphery, and furthermore that the occlusion only blocks S's view of Susan's body, without blocking the view of her face. Let us call this a case of *Insignificantly occluded seeing*. If the process is typed as *Insignificantly occluded seeing*, we may once more infer reliability. And so things could continue.[26]

[26] As Goldman reminded me, Brandom (1998) describes analogously unending switches in reliability as one moves from one process category to another.

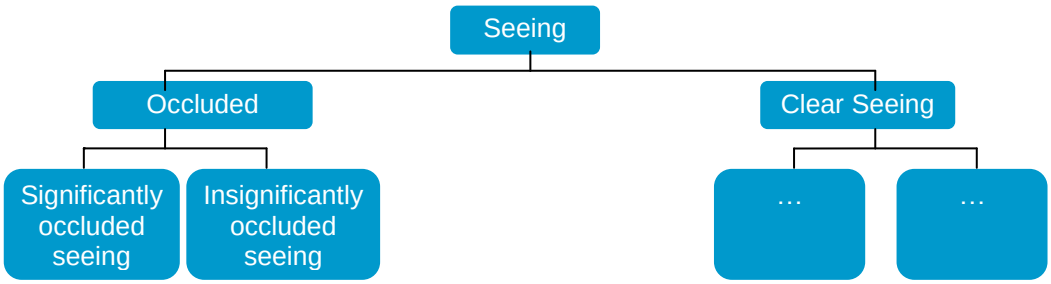


Figure 3: A hypothetical non-monotonic taxonomy

If we focus on the single purpose classificatory task, the attribute of reliability/unreliability would be assigned as follows:

Seeing: reliable

Occluded seeing: unreliable

Insignificantly occluded seeing: reliable

Significantly occluded seeing: unreliable

One can imagine how a taxonomy like the one in Figure 3 could be generated "on the fly", as in the following dialogue:

John: I know Susan is over there. I see her.

Mary: How can you be sure given the occlusion?

John: True, but I see her face so the occlusion is insignificant.[27]

Logically speaking, nothing prevents ever more fine-grained types from being proposed endlessly, but that would surely be quite unrealistic in practice. Rather, we would expect the process of generating types on the fly soon to come to an end. At that point, there are a finite number of types "on the table" forming a taxonomy

[27] Non-monotonic effects arise in scientific taxonomies as well. In biology, the category *Whale* is subordinate to the category *Mammal* and yet the mammal attribute "lives on land" is not inherited by the *Whale* category. Another classic example involves penguins, i.e., birds that, unlike most, do not fly.

of types. The question then is: Is there a basic level even if the taxonomy is non-monotonic and, if so, what characterizes that level?

The Rosch rule for single purpose categorization, as I have construed it, states that the basic level is the most inclusive level in the taxonomy at which reliability or unreliability would be listed. If the occlusion was in fact insignificant, this means that the Rosch rule singles out the category of *Seeing* as the basic level category, from which we may infer that the process was, as we would have expected it to be, reliable. The problem is that the Rosch rule gives exactly the same result if we assume instead that the process was one of significantly occluded seeing. In that case, too, we get the result that *Seeing* is the basic level category so that the process was reliable, which is simply incorrect. Thus, while the Rosch rule for single purpose categorization is plausible for monotonic taxonomies, it fails to account for basic level effects in the presence of non-monotonicity.

An alternative suggestion comes to mind: to focus initially on categories to which the process belongs that cannot be further subdivided in ways that would make a difference to what can be inferred about the reliability or unreliability of the process in question; and to choose, among the candidates that remain, the category that is most inclusive. Suppose again that the occlusion was insignificant, which means that the token process belongs to three categories: *Seeing*, *Occluded seeing* and *Insignificantly occluded seeing*. Only the latter cannot be further subdivided in a way that is relevant to reliability, which means that it will be singled out as the process type, enabling us to infer that the process was reliable. By the same token, if the occlusion was significant, the process will be typed as *Significantly occluded seeing*, from which we may infer that it was unreliable.[28]

Unfortunately, as Goldman has pointed out to me (personal communication) this outcome is problematic because the selected categories will, as the example illustrates, be more specific than the expected commonsense categories, i.e., *Seeing*, *Hearing* etc. Another way of putting it is that the proposal is difficult to

[28] This proposal bears some similarity, at an abstract level, to the statistical rule advocated in Beebe (2004).

justify from a Roschian perspective. Here is why: As we recall, two cognitive goals are taken, within that framework, to be universally operative in human cognition: the goal of economy and the goal of usefulness. It could be agued that, while the present proposal may be justifiable from the point of view of usefulness, it does not do justice to the goal of cognitive economy. Let us focus on the latter point. As we saw, a process that is a case of insignificantly occluded seeing is assigned the type *Insignificantly occluded seeing* by the proposed rule. Based on that classification, it can be inferred that the process was reliable. But the same inference could have been made given a coarser and therefore more economical typing of the process as *Seeing*. Similarly, a process that is a case of significantly occluded seeing is assigned the type *Significantly occluded seeing*, which allows an inference to unreliability. Once more, the same inference could also have been drawn by more economical means by typing the process as *Occluded seeing*. The proposed rule is, for these reasons, clearly sub-optimal from the standpoint of the two cognitive forces that are assumed by Rosch et al to be operational in categorization.

In response to this concern, which Goldman has described in our personal correspondence as his "principal objection" to the present approach, I will propose an amended rule that improves on the previous proposal, drawing on the insights we just acquired as to why the latter sometimes gives suspect results. Let us say that the single purpose category for token process t for taxonomy T (with respect to reliability) is the most inclusive T-category C such that (i) t belongs to C and (ii) C is associated with the same reliability attribute (reliable/unreliable) as is the most specific T-category to which t belongs. In other words, we first check what can be inferred regarding reliability from the most specific correct categorization of a given token process in the taxonomy and then identify the most inclusive superordinate from which that same inference can be drawn. As can easily be checked, the amended rule will fare better than the previous proposal with respect to economy, without any informational loss being incurred, because it will type insignificantly occluded seeing as *Seeing* and significantly occluded seeing as *Occluded seeing*. I take it that this modified rule thereby also answers Goldman's

objection. For monotonic taxonomies the modified rule gives the same result as the original Rosch rule.

5. Discussion and conclusion

This completes my account of why we should, from the standpoint of cognitive science, expect people to agree to a significant extent in their typing of belief forming processes and why, as a result, they should also agree on the resulting reliability verdict. My aim has not been to provide a complete psychological story for why convergence should be expected. I believe, however, that it suffices for the purposes at hand: enough has been said to shift the burden of proof to Conee and Feldman, who, as we recall, predict massive category II disagreement on reliability in the absence of linguistic guidance, a contention for which they offer little or no evidence beside their own intuitions. Hence, I take what has been said above to be a sufficient response to the generality problem as an objection that specifically targets reliabilism.

To recapitulate, the reasons provided by Conee and Feldman for thinking that the reliabilist analyses of knowledge and justification lack sufficient similarity to the corresponding everyday concepts could be neutralized by reference to influential work in cognitive psychology. Now the way in which cognitive psychology solves this problem is surprisingly consonant with Conee and Feldman's thinking about the criteria that an acceptable solution to the generality problem should satisfy. Typing takes place in a cognitive environment featuring two contradictory cognitive forces of economy and usefulness. In an effort to strike a balance between these two forces, the cognitive subject will select a category that is reasonably economic while preserving as much potentially useful information as possible. The result will be a middle range category. This shows that the present account not only avoids the single case and no distinction problems; it also accords with the Conee and Feldman condition that an acceptable account should be "principled". Furthermore, the solution clearly

meets their further criterion of being in the spirit of reliabilism, especially as the latter is advanced by practitioners of naturalized epistemology.

What about the remaining criterion stressed by Conee and Feldman – that the types identified must have a reliability that is plausibly correlated with the justificational status of the beliefs in question? I see no reason to think that this condition should not be satisfied as well. The present theory predicts that people will normally type processes, even in the absence of linguistic hints, as *Seeing*, *Hearing* etc. It should be expected that the perceived reliability of these processes will be strongly correlated with the perceived degree of justification of the resulting beliefs. In fact, recent experiments carried out by my colleague Martin L. Jönsson along the lines described in section 2 amply support this prediction (Jönsson, forthcoming). Jönsson found that there to be significant agreement among the folk about how to type belief forming processes. For some stimuli, the agreement was even 100 percent. He also discovered there to be a strong correlation between the perceived degree of reliability and the perceived degree of justification.[29]

For the purposes of neutralizing Conee and Feldman's central challenge it was sufficient to give a purely descriptive account of how people type belief forming processes. To be sure, such an account does not give us a method for selecting the normatively correct type in a given case. A normative method of this kind would indeed be a good thing to have, especially in cases in which there are conflicting views about how a given process should best be typed, although it is worth repeating that it is no targeted objection to reliabilism to point out that such a method is still missing. A further thought is that the Rosch theory may actually turn out to be helpful in this connection as providing a theoretical framework

[29] A further worry raised by Goldman in our personal communications centered on the issue whether the agreement observed in people's responses may be subject to instability, e.g. dependent on the order in which the examples were presented, analogously to how Swain, Alexander and Weinberg (2008) found the order of examples to be significant in how people judged Gettier cases. In order to eliminate any doubts of this nature, Jönsson was careful to construct his experiment so as to eliminate any dependence of the results on the order in which examples were presented. The examples used by Jönsson involved everyday scenarios rather than Gettier cases. Further experimentation would be needed in order to ascertain whether people disagree about typing and reliability in Gettier cases, or whether their intuitions in such scenarios about these things are unstable (order dependent etc.).

within which a normative account could be developed, perhaps by epistemologists and categorization theorists working together.

As we saw, finally, Goldman's principal objection prompted me to think more carefully about the non-monotonic case, which in turn led up to a slightly revised account of the Rosch rule for categorization which, in my view, solves the problem in conformity with the most entrenched parts of the Rosch theory. The result is a response to what is to my knowledge the only version that should prompt the reliabilist's serious attention, a response which Goldman as far as I can see could adopt without sacrificing any essential part of his epistemological framework.

Acknowledgements

This paper is a contribution to the research project *Knowledge and Reliability* funded by the Swedish Research Council. Many people have commented on various versions of the paper, and I am grateful to all of them for their input. They will know who they are. I have also benefited from the criticism of various seminar and conference audiences, including the participants of the higher seminar in theoretical philosophy at my department in at Lund University in Sweden and the members of the Copenhagen-based SERG research group, headed by Klemens Kappel. I am above all deeply grateful to Alvin Goldman for generously sharing his insights on epistemology in general, and reliabilism in particular, on many occasions over the years we have known each other; for commenting extensively on my previous attempts to come to grips with the generality problem, which massively help me in sharpening my thinking; and, finally, for allowing me to draw on these personal correspondences when writing this paper.

References

Adler, J., and Levin, M. (2002) Is the Generality Problem too General? Philosophy and Phenomenological Research 65(1), 87-97.

Alston, W. (1995) How to Think about Reliability. Philosophical Topics (Spring), 1-29.

Beebe, J. R. (2004) The Generality Problem, Statistical Relevance and the Tri-Level Hypothesis, Nous 38(1), 177-195.

Brandom, R. B. (1998) Insights and Blindspots of Reliabilism. The Monist 81(3), 371-392.

Carnap, R. (1950) Logical Foundations of Probability. Chicago University Press: Chicago.

Christensen, D. (2007) Three Questions about Leplin's Reliabilism. Philosophical Studies 134(1), 43-50.

Comesaña, J. (2006) A Well-founded Solution to the Generality Problem. Philosophical Studies 129, 27-47.

Conee, E. and Feldman, R. (1998) The Generality Problem for Reliabilism. Philosophical Studies 89, 1-29.

Corter, J. E., and Gluck, M. A. (1992) Explaining Basic Categories: Feature Predictability and Information. Psychological Bulletin 111(2), 291-303.

Feldman, R. (1985) Reliability and Justification. The Monist 68, 159-174.

Feldman, R., and Conee, E. (2002) Typing Problems. Philosophy and Phenomenological Research 65(1), 98-105.

Goldman, A. I. (1979) What is Justified Belief? In: Pappas, G. (ed.) Justification and Knowledge. D. Reidel, Dordrecht, pp. 1-23.

Goldman, A. I. (1986) Epistemology and Cognition. Harvard University Press, Cambridge, MA.

Goldman, A. I., (2008) Reliabilism. In Zalta, E. N. (ed.) The Stanford Encyclopedia of Philosophy, URL = < http://plato.stanford.edu/entries/reliabilism/>

Goldman, A. I., and Olsson, E. J. (2009) Reliabilism and the Value of Knowledge. In Haddock, A., Millar, A. and, Pritchard, D. H. (eds.) Epistemic Value, Oxford University Press, pp. 19-41.

Heller, M. (1995) The Simple Solution to the Problem of Generality. Noûs 29(4), 501-515.

Hetherington, S. C. (1996) Knowledge Puzzles: An Introduction to Epistemology. Westview, Boulder, CO.

Jönsson, M. L. (forthcoming) A Reliabilism Built on Basic Levels.

Jönsson, M. L., and Hampton, J. A. (2006) On Prototypes as Defaults. Cognition 106(2), 913-923.

Jönsson, M. L., and Hampton, J. A. (2008) The Inverse Conjunction Fallacy. Journal of Memory and Language 55(3), 317-334.

Kappel, K. (2006) A Diagnosis and Resolution to the Generality Problem. Philosophical Studies 127, 525-560.

Lemos, N. (2007) An Introduction to the Theory of Knowledge. Cambridge University Press, New York.

Leplin, J. (2007) In Defense of Reliabilism. Philosophical Studies 134(1), 31-42.

Lycan, W. G. (1988) Judgment and Justification. Cambridge University Press, New York.

Olsson, E. J. (2004) F. P. Ramsey on Knowledge and Fallibilism. Dialectica 58(4), 549-557.

Olsson, E. J. (2007) Reliabilism, Stability, and the Value of Knowledge. American Philosophical Quarterly 44(4), 343-355.

Olsson, E. J. (2008) Knowledge, Truth, and Bullshit: Reflections on Frankfurt. Midwest Studies in Philosophy 32, 94-110.

Olsson, E. J. (2009), In Defense of the Conditional Probability Solution to the Swamping Problem, Grazer Philosophische Studien 79, 93-114.

Olsson, E. J. (2011) Reply to Kvanvig on the Swamping Problem. Social Epistemology 25(2), 173-182.

Olsson, E. J.. and Jönsson, M. L. (2011) Kinds of Learning and the Likelihood of Future True Beliefs: Reply to Jäger on Reliabilism and the Value Problem. Theoria 77(3), 214-222.

Olsson, E. J. (in press) The Generality Problem Naturalized. In: Svensson, F., and Sliwinski, R. (eds.) Festschrift for Erik Carlson on the occasion of his 50th birthday. Uppsala Philosophical Studies, Uppsala.

Plantinga, A. (1993) Warrant: The Current Debate. Oxford University Press, Oxford.

Pollock, J. L. (1986) Contemporary Theories of Knowledge, Rowman and Littlefield, Totowa, NJ.

Pollock, J. L., and Cruz, J. (1999) Contemporary Theories of Knowledge, 2nd edn. Rowman and Littlefield, Lanham, MD.

Ramsey, F. P. (1931) The Foundations of Mathematics and other Logical Essays. Braithwaite, R. B. (ed.), Routledge and Kegan Paul, London.

Rogers, T. T., and Patterson, K. (2007) Object Categorization: Reversals and Explanations of the Basic Level Advantage. Journal of Experimental Psychology: General 136(3), 451-469.

Rosch, E,. Mervis, C. B, Gray, W. D, Johnson, D. M, and Boyes-Braem, P. (1976) Basic Objects in Natural Categories. Cognitive Psychology 8, 382-439.

Swain, S., Alexander, J., and Weinberg, J. (2008) The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp. Philosophy and Phenomenological Research 126(1), 138-155.

Wunderlich, M. (2003) Vector Reliability: A New Approach to Epistemic Justification, Synthese 136, 237-262.

Zacks, J. M., and Tversky, B. (2001) Event Structure in Perception and Conception. Psychological Bulletin 127(1), 3-21.