Green is Simpler than Grue *

Itzhak Gilboa[†]

October 2007

Abstract

Goodman's grue-bleen paradox relies on the claim that the predicate "green" is not simpler than the predicate "grue", and that, for that reason, the choice between them is arbitrary. It is argued that this claim can only be made in a restricted model, in which functions have primitive names but particular values do not. In more natural models, which allow values to have proper names as well, the paradox disappears.

1 Goodman's grue-bleen paradox

Goodman's grue-bleen paradox (Goodman, 1955) can be described as follows.

Assume that a scientist wishes to test the theory that emeralds are green, contrasted with the theory that they are blue. Testing one emerald after the other, she concludes that emeralds are indeed green.

Next assume that another scientist comes along, and wants to test whether emeralds are grue as opposed to bleen. "Grue" emeralds are emeralds that appear to our eyes green if tested until time T, but appear blue if they are tested after time T. "Bleen" emeralds are defined symmetrically. Choose a time T in the future, and observe that the scientist will find all emeralds to

*I wish to thank Orna Harary, Philippe Mongin, Dov Samet, David Schmeidler, and Teddy Seidenfeld for comments. No endorsement of content should be inferred from this acknowledgment.

[†]Tel-Aviv University, HEC, and Cowles Foundation, Yale University. igilboa@tau.ac.il

be grue. She will therefore conclude that after time T, all new emeralds to be tested, which will most likely be grue as those tested up to T, will appear blue to our eyes.

Testing the hypotheses that emeralds are green versus blue seems perfectly reasonable. Testing for grueness versus bleenness appears weird at best. Yet, how can we explain our different reactions? The two procedures seem to be following the same logical structure. They both appear to be what the scientific method suggests. Why do we accept one and reject the other?

A common reaction is to say that, if emeralds are tested and found green, there is no reason to suppose that, all of the sudden, new emeralds will start appearing blue. But one may claim, by the same token, that there is no reason for new emeralds to switch, at time t, and become bleen. Our common sense suggests that "green" and "blue" are simple predicates, whereas "grue" and "bleen" are complex predicates. But this is true only if one assumes "green" and "blue" as primitives of the language, and defines "grue" and "bleen" in terms of "green" and "blue". If one were to start out with "grue" and "bleen" as the primitive terms, "green" and "blue" would appear to be the complex terms.

In this note we try to argue that the preference for simplicity can indeed explain the preference for the generalization by the predicate "green" over that by "grue". We take for granted a preference for simplicity, as in

> "The procedure of induction consists in accepting as true the *simplest* law that can be reconciled with our experiences." (Wittgenstein, 1922, 6.363).

Such a preference for simplicity may be justified on evolutionary grounds,[1] but this is not the focus on this note. It will also be assumed that we tend

[1] One simple argument is that an unboundedly rational species that has no preference for simpler theories is bound to "overfit" the data and therefore to be a poor learner.

to choose a language based on simplicity considerations, namely one within which simple theories can explain the data.[2] The question is, can we justify the preference for the predicate "green" over "grue" *assuming* a preference for simplicity.

We provide two related but separate arguments. One is an informal evolutionary parable. The other is a formal model. Both make a similar point, namely, that Goodman's paradox relies on a very peculiar choice of language.[3]

2 An Evolutionary Parable

One might try to imagine a process by which a very "low" form of life starts evolving into a smart human. Suppose that the animal in question has eyes, and its eyes reflect the image they see in some neurons. The animal may have a certain neuronal firing configuration when it sees the color green (i.e., receives wave lengths that correspond to what we call green in everyday life) and others – for blue. These neuronal configurations will probably be rather stable, so that a particular configuration would be correlated with particular wave lengths whenever we (as outside observers) perform the experiment. Importantly, such a stable correlation can be established way before the animal has any conception of time, past and future, or any ability to generalize. That is, *we* know that "for every time t, configuration c is associated with wavelength w", but this knowledge cannot be attributed to the animal. The animal cannot reason about t.

We can go further and endow the animal with the ability to speak and to name the colors, still without allowing it to think about time. That is, our creature observes a wavelength w, has a neuronal reaction c, and may learn

[2] The preference for a "simple language" can be reduced to the preference for a simple theory. Specifically, if one models theories as computer programs, and their complexity – by their length (Kolmogorov complexity), the choice of a language may be viewed as the choice of "procedures" defined in a program.

[3] I am actually not convinced that the evolutionary parable adds much. Formally oriented readers may skip it.

to make an utterance u, naming the color as "green" or "blue". Assuming that naming colors is useful (say, for communication and coordination with others), the creature can say the correct color word whenever presented with the color. We can go as far as to say that the creature *knows* what color is presented to it. We can make this generalization and argue that, at every time t, the creature knows what is the color presented. Still, no quantifier over t can be ascribed to our creature, which is still ignorant of time and incapable of generalizations.

At the risk of repeating the obvious, let us stress that the "green" or "blue" uttered by our creature are not the same "green" or "blue" to which Goodman refers. Goodman thinks of "green" as a property that involves generalization and induction. That is, for him "green" means "for every time t, if tested at t, then appears green to our eyes".[4] Not so for our creature. The latter cannot think in terms of "for every time t", not even in terms of "now". We, the outside observers, know that the creature responds, at time t, to the input presented at time t. But there is no clock inside the creature's mind, and there is no consciousness of time.

Assume now that another layer of intelligence is added to our creature, and the notion that it is going to live also in the future finally dawns on it. Starting to think about time, it may ask itself what is the nature of emeralds. This "nature of emeralds" will now become a function from time into the two timeless colors, i.e., the neuronal configurations "green" or "blue". Again, these are not the predicates "appears green whenever tested" or "appears blue whenever tested"; these are simply the names that the creature learnt to associate with certain neuronal configurations. But if such names already existed in its language, the functions "appears green whenever tested" or

[4] Or, "for every time t, if tested at t, emanates waves in the green range". We do not mean to raise a new question about the relationship between wavelengths and subjective color experience. The point is only to refer to the specific experience that occurs at a given time, without generalizing it to a phenomenon that generates the same experience at every t.

"appears blue whenever tested" seem simpler than the functions describing "grue" and "bleen".

3 A More Formal Treatment

3.1 Re-stating Goodman's paradox

The formal model we use here is standard and straightforward. But to see exactly where Goodman's paradox arises from, one may need to develop the model with care that might seem pedantic.

Denote time periods by $\mathbb{N} = \{1, 2, 3, ...\}$. At each time i there is a value $x_i \in \{0, 1\}$, which has been observed for $i \leq T$, and is to be predicted for $i > T$. The value x_i should be thought of as the color of the emerald tested at period i. For example, $x_i = 0$ may denote the observation of an emerald that appeared blue (i.e., in the blue wavelength), and $x_i = 1$ – that appeared green. The model can be extended to an arbitrary set of observations at no additional cost. As usual, finite sets can be encoded as sequences of binary digits, so that the present formulation does not involve any significant loss of generality. It is important, however, that the present analysis does not necessitate additional observations or embedding the problem in a larger context.[5]

Let $F = \{0, 1\}^{\mathbb{N}}$, the set of functions from time periods to observations, denote all possible theories. We focus here on generalizations from the past to the future, but the same apparatus can be used to analyze any projection problem, that is, any generalization from a set of observations to a superset thereof.[6]

[5] Some resolutions of Goodman's paradox do rely on a wider context, such as Goodman's own theory of entrenchment. See the discussion below.

[6] Goodman was, of course, correct to point our that a prediction problem can be viewed as special case of a projection problem. At the same time, the opposite embedding can also be done: if we enumerate observations by the time at which they were revealed to us, a projection problem can be thought of as the problem of predicting what our next observation will be, even if we believe that the content of the observation has been determined

It will be useful to re-state the problems of induction explicitly within the formal model.

(i) How can we justify generalization: Suppose that a function $f \in F$ matches all past data, that is, $f(i) = x_i$ for all $i \leq T$. How do we know that f will continue to match the data for $i > T$?

(ii) How do/should we perform generalizations: Suppose that $f, g \in F$ are such that $f(i) = g(i) = x_i$ for all $i \leq T$, but $f(i) \neq g(i)$ for some $i > T$. How do we know whether f or g should be used for prediction?

(iii) What is a natural generalization: In the set-up above, which of the two theories may qualify as a "simpler" theory or otherwise a "natural" choice?

Hume (1748) famously posed the first problem and argued that no logical justification of generalization is possible. This has been recognized as a problem we have to live with.[7] But once we agree that there is no logical necessity that a function f will continue to match the data in the future, we are faced with the second problem, namely, that there are many possible ways to project past observations into the future. Which one should we choose?

One may resort to simplicity as a justification for inductive reasoning. Simplicity provides a criterion for selection of a theory f among all those that fit past observations. Indeed, everyday reasoning and common scientific practice implicitly deal with problems such as (ii), and statistical inference does so explicitly. It appears that simplicity, again, is the key to induction, as suggested by Wittgenstein in the quote above. It is an insightful description of our natural tendency to choose theories in everyday and in scientific reasoning, and it is a prominent principle for model selection in statistics.[8]

in the past. In this sense, historians can also engage in prediction, say, speculating what a new excavation in an archeological site will reveal.

[7] Quine (1969) writes, "... I do not see that we are further along today than where Hume left us. The Humean predicament is the human predicament".

[8] Statistical theory assumes that the data are generated by a process that is, as far as we can tell, inherently random. Hence none of its models fit the data perfectly. Instead, the theory explicitly copes with the trade-off between a model's accuracy and its simplicity.

Simplicity does not always offer an unique theory that is clearly the obvious candidate to generalize the observations. For instance, given the observations 01, one may be unable to make a reasoned choice between the competing generalizations 011111... and 010101... . Yet, each of these latter sequences offers a natural continuation (as is evident from the informal way in which we described them in the previous sentence). On the other hand, a sequence such as 011010 allows, again, for a variety of generalizations.

It follows that, if we accept the preference for simplicity, Problem (ii) can be easily solved as long as, among all theories that conform to observations, there is one that is much simpler than all the others. But Problem (ii) will remain a serious problem if the data do not allow such a theory to emerge. Indeed, when the data appear "random",[9] the simplest theories that conform to the data are going to be many equally complex theories. Generally speaking, our confidence in a generalization will be a matter of degree, depending on quantifiable magnitudes such as the randomness in the data, the theory's simplicity, the simplicity of competing theories, and so forth.

Given this background, we can finally consider Goodman's paradox. Goodman posed problem (iii), suggesting that it is not always obvious what it the most natural generalization, or which is the simplest theory that fits the data. This problem is not always a "paradox". It is often a very real problem, and serious experts may indeed debate how should past observations be generalized. However, in Goodman's emeralds example problem (iii) appears paradoxical: we all feel that "green" is a natural generalization whereas "grue" is a ridiculous one, but we are hard-pressed to explain why. In the rest of this section we wish to explain why problem (iii) is not a serious problem in the case of the emeralds, but we do not wish to dismiss it altogether. To re-iterate, we find that this problem is a very important one in general,

[9]The term "random" is used here informally. Kolmogorov turned this argument around and used the minimal length of a theory tha generates the data as a measure of complexity, and defined randomness of a sequence by this measure.

and there is nothing paradoxical or perplexing about its very existence.[10] At the same time, we need to explain why it seems silly when emeralds are concerned.

Goodman considers emeralds that have so far been observed to be green. That is, he considers a long sequence 111... . According to the discussion above, as well as to our intuition, there is but one reasonable way to extend this sequence, namely, to consider the function $f_{green} : \mathbb{N} \to \{0,1\}$ that is the constant 1, namely,

$$f_{green}(i) = 1 \qquad \forall i \in \mathbb{N} \tag{1}$$

function, . Goodman contrasts this function with another function $f_{grue} : \mathbb{N} \to \{0,1\}$ defined by

$$f_{grue}(i) = \begin{cases} 1 & i \leq T \\ 0 & i > T \end{cases} . \tag{2}$$

The standard argument is that we should prefer f_{green} over f_{grue} on the basis of simplicity. If we were, say, to describe f_{green} and f_{grue} by PASCAL programs, f_{green} will have a shorter description than will f_{grue}.[11] But here Goodman asks that we realize that one need not assume "green" and "blue" as primitives of the language. If, instead, we use "grue" and "bleen" as primitives, we find that f_{grue} is "grue whenever tested", whereas f_{green} is given by "grue if tested at time $i \leq T$ and bleen otherwise", in which case "green" and "blue" are the complicated functions.

Stating Goodman's argument formally would highlight the problem. Define also the functions f_{blue}, and f_{bleen} in the symmetric way, namely,

$$f_{blue}(i) = 0 \qquad \forall i \in \mathbb{N} \tag{3}$$

[10] See the discussion below.

[11] PASCAL is a computer language that was originally developed as a tool to describe algorithms for formal purposes. It is a much more natural language to describe algorithms than is the language of Turing machines. However, when applied to a computer with an infinite memory, these languages, as well as many others, are equivalent in terms of their descriptive power.

and

$$f_{bleen}(i) = \begin{cases} 0 & i \leq T \\ 1 & i > T \end{cases} . \tag{4}$$

Goodman suggests that we consider the definitions of f_{grue} and f_{bleen} in the language of f_{green} and f_{blue}, i.e.

$$f_{grue}(i) = \begin{cases} f_{green}(i) & i \leq T \\ f_{blue}(i) & i > T \end{cases} \qquad f_{bleen}(i) = \begin{cases} f_{blue}(i) & i \leq T \\ f_{green}(i) & i > T \end{cases} \tag{5}$$

and contrast them with the definitions of f_{green} and f_{blue} in the language of f_{grue} and f_{bleen}, namely

$$f_{green}(i) = \begin{cases} f_{grue}(i) & i \leq T \\ f_{bleen}(i) & i > T \end{cases} \qquad f_{blue}(i) = \begin{cases} f_{bleen}(i) & i \leq T \\ f_{grue}(i) & i > T \end{cases} \tag{6}$$

Goodman is absolutely right to argue that (5) is just as complex as (6). But his formulation does not allow us to compare the complexity of (2) and (4) with that of (1) and (3).

In other words, Goodman refers to the pairs of predicates "green whenever tested" and "blue whenever tested", vs. "grue whenever tested" and "bleen whenever tested", and argues that the choice among them is arbitrary. That is, he refers to functions in F. But he does not refer to the values that these functions assume, 0 and 1. When he refers to an observation of an emerald as green, he does not treat it as observing the value 1, but as observing that at time i the data were consistent with a certain theory f. That is, he claims that we observed $f(i)$ (for some f). He does not allow us to say that we observed 1.

Of course, whenever $f(i) = 1$ for some function (predicate) f, one may replace 1 by $f(i)$ and obtain a valid description of the observations. But one should not be surprised if complexity considerations do change as a result of this substitution. Indeed, one lesson we learnt from Goodman's paradox is that the measurement of complexity depends on language. In a natural language, where the values 0 and 1 have explicit references, or proper names, the function f_{green} above is simpler than f_{grue}. But if we are not allowed to use names for the specific values, and we can only refer to them indirectly as the values assumed by generalized functions, then indeed f_{green} and f_{grue} may appear equally simple. In other words, Goodman's paradox relies on a very peculiar language, in which there are names for functions but not for the specific values they assume.

3.2 Another example

At the risk of boring the reader, let us consider another example of the Goodman type. The key to constructing such examples is that we consider formal languages whose only primitives are functions $f : \mathbb{N} \to \{0, 1\}$.

Consider the following two pairs of functions: f_0 and f_1, which are the constant functions equal to 0 and to 1, respectively; and the pair g_{odd} and g_{even}, where $g_{odd}(i) = 1$ iff i is odd, and vice versa for g_{even}. Again, these functions are the *only* primitives of the language. That is, if we wish to say that $1 - 1 = 0$, we may write

$$f_1(2) - f_1(7) = f_0(15)$$

or

$$g_{odd}(9) - g_{even}(8) = g_{even}(7)$$

but no simpler expressions exist. That is, the only way to refer to a number, 0 or 1, is as a specific value of one of the general functions for a given time period i.

If we can state logical conditions in terms of the time period (the argument of the functions discussed), we find that the language of $\{f_0, f_1\}$ has the same expressive power as that of $\{g_{odd}, g_{even}\}$. Explicitly,

$$g_{odd}(i) = \begin{cases} f_1(i) & i \text{ odd} \\ f_0(i) & i \text{ even} \end{cases}$$

and so forth.[12] Thus, the choice the language – $\{f_0, f_1\}$ vs. $\{g_{odd}, g_{even}\}$ – should be determined by the simplicity of the theories that conform to observations, when stated in these languages. If, for instance, we observe the sequence 10101010, the theory "g_{odd} for every i" would appear to be a simple theory that matches the data very well, whereas explaining the same observations by the theory that switches between f_1 and f_0 would appear cumbersome. However, if we observe the sequence 110011001100, explanations in terms of $\{f_0, f_1\}$ and in terms of $\{g_{odd}, g_{even}\}$ are equally cumbersome. In this case, there is no reason to prefer one language over the other. Most of us still feel that $\{f_0, f_1\}$ are simpler and more natural primitives than are $\{g_{odd}, g_{even}\}$. But in the context described above there is no way to capture this alleged simplicity. Describing $\{f_0, f_1\}$ in terms of $\{g_{odd}, g_{even}\}$ is just as simple as describing $\{g_{odd}, g_{even}\}$ in terms of $\{f_0, f_1\}$.

This example is akin to Goodman's paradox. Both "green" and "grue" describe past observations just as well. This is similar to descriptions of a sequence such as 110011001100 in both languages. We have a strong sense that "green" is a simpler predicate than "grue", just as we feel that $\{f_0, f_1\}$ are simpler than $\{g_{odd}, g_{even}\}$. But if the language does not refer to specific values of these functions, it is hard to justify the simplicity judgments above.

On the other hand, the problem is resolved as soon as our formal language allows symbols such as 0 and 1. Given a sequence such as 10101010, we may choose to use the term g_{odd}, and, if needed, to define it in terms of the basic observations 0 and 1 so as to facilitate future discussion. Yet, g_{odd} is a-priori less natural than f_0 or f_1, and g_{odd} will be the term of choice only if it fits the data better than simpler terms.

[12] Observe that the explicit reference here is only to the argument i and not to the values assumed by the functions.

3.3 Should values have names?

There are several reasons for which one may require that our language have names for specific values, namely, for observations. The first and perhaps most compelling is, why not? A true and fundamental problem cannot be alleviated by the introduction of terms whose observational content can be defined by existing ones.

To see this point more clearly, observe that one may define values by using variables. For instance, given that the function f_{grue} (or f_{green}) is in the language, one would like to be able to set

$$x = f_{grue}(1); \qquad y = 1 - f_{grue}(1)$$

and thereafter use also the variables x and y in the description of theories. This would be a rather awkward substitute to using the values 0 and 1, but even this would not allow us to state the paradox. The paradox relies on a ban on references to observations either as specific values or as variables assuming such values.

A second reason to include names for values is that they would enable the modeling of inductive reasoning, namely the process by which particular sense data are generalized to predicates such as "green whenever tested" and "blue whenever tested".

This point may be worth elaboration. Goodman's formulation actually does not allow us to say "the emerald tested at time t was green". We can only say, "the emerald tested at time t was of the type green-whenever-tested". The fact that this is not obvious is precisely because we don't expect to see grue emeralds. Since all emeralds so far tested were found to be green, we perform the inductive step and thus tend to confuse "green" with "of the type green-whenever-tested", that is, the specific value and its generalization.

Suppose, by contrast, that we try to forecast the weather. Every day we may observe rain or no rain. If it rained on day t, we may say that $x_t = 1$, or even that day t was "rainy". But we do not say "day t was of the type rainy-at-each-i". Because rain is not as regular a phenomenon as the color of emeralds, it would be unlikely to confuse "my hair is getting wet" with "this is an instance of the theory that my hair always gets wet".

To conclude, if we take a simplistic point of view, according to which observations precede and exist independently of theories, it is natural to require that our language be able to describe the data themselves, without engaging in theorizing about them. To discuss theory-free observations, one needs to be able to refer to specific values, that is, to observations. But even if one takes a more realistic point of view, according to which observations are theory-laden (Hanson, 1958), it is hard to come up with a natural example in which observations do not have names in the language.

4 Concluding Remarks

4.1 Complexity and language

The recurrent reference to simplicity raises questions on the relationship between simplicity and language. The formal treatment above argued that the constant function f_{green} was simpler than the function f_{grue}. But why is f_{green} simpler than f_{grue}? The notion of Kolmogorov's complexity was mentioned in this context, suggesting that the (minimal length) description of f_{green} as a PASCAL program will be shorter than that of f_{grue}. This, however, raises the question, is PASCAL the most natural language to use? Maybe there exists another language, equally attractive as is PASCAL, in which f_{grue} will appear simpler than f_{green}, and the paradox will be reinstated?

It is important to emphasize that the measurement of complexity does indeed depend on the language we use. Goodman should probably be credited as the first author to make this point.[13] When measures of complexity were formally developed by Kolmogorov (1963, 1965) and Chaitin (1966), and also applied to philosophy of science by Solomonoff (1964), this point

[13] See also Sober (1975), who quotes Goodman on this point.

was explicitly discussed. Indeed, when undertaking a formal analysis, one can hardly avoid the question of the choice of language when one attempts to write down a well-defined model. Having said that, two points are in order.

1. The power of Goodman's paradox is not in the general claim that language affects simplicity judgments, but in the concrete example of the predicates green and grue. It is perplexing because Goodman shows that the natural predicate is just as arbitrary as the counter-intuitive one. The point of this note is that this equivalence relies itself on a particular choice of a restricted and unnatural language.

2. The general claim that simplicity judgments depend on language is qualified by the observation that languages with the same expressive power have compilers translating from one to the other. Specifically, for two languages l and l', let $C_{l,l'}$ be a program, written in l', that accepts as input a program written in l and generates as output an equivalent program written in l'. Let $c_{l,l'}$ be the length of the minimal description of such a compiler $C_{l,l'}$. It follows that, if the minimal length of a program (representing a theory) t in l is m, the minimal representation of t in l' cannot exceed $m + c_{l,l'}$. Hence, for every two languages l and l', there exists a constant $c = \max(c_{l,l'}, c_{l',l})$ such that, for every theory t, the complexity of t as measured relative to l and the complexity of t as measured relative to l' cannot differ by more than c.

The order of quantifiers is of importance here: given two theories t and t', one can always come up with a language l, in which one theory is simpler than the other. On the other hand, given any two languages l and l', one finds that, if theory t is sufficiently simpler than theory t' in one language, t has to be simpler than t' also in the other language.

Specifically, consider the complexity of f_{green} and f_{grue}, as defined in (1,2) in PASCAL. The complexity of the theory f_{green} is finite. By contrast, the complexity of f_{grue} grows with T. Consider any other language in which the theories can be described. By the reasoning above, for a large enough T that language will also rank f_{grue} as more complex than f_{green}.

4.2 Other resolutions

Much has been written about Goodman's paradox, and the present note need not contradict any other explanation of the paradox. In particular, Goodman's own theory of entrenchment adds to our understanding of the nature of induction. However, there is no need to embed the problem in a wider set-up in order to resolve the paradox.

There have been claims that "grue" is not a testable predicate whereas "green" is. In a sense, there is truth to this claim, though it might be misleading without a careful analysis. If "green" and "grue" are understood as the general functions, "green whenever tested" and "grue whenever tested", then they are equally testable. Observing a green emerald at time $i \leq T$ can be described as "observing an emerald of the type green-whenever-tested" or as "observing an emerald of the type grue-whenever-tested", both confirming the respective theories but obviously not proving them. However, if we restrict attention to specific values, then we find that "green" is a name of such a value, whereas "grue" is not. Again, the confusion between the two meanings of "green" is the heart of the problem: as a name of a value, it has no counterpart "grue"; as a name of function, it does. The functions, as general theories, are equally testable. As a specific value, only "green" exists.

4.3 Lessons for the definition of probabilities

This discussion is not merely of philosophical interest. The definition of probabilities in everyday life often encounters problems of generalization that are similar in spirit to Goodman's paradox. The basic point, namely, that there may be more than one way to generalize the past to the future, is one of the reasons that empirical frequencies are inadequate for the definition of probability for many events of interest. In examples such as green and blue, where there is a great deal of stability in our observations, the paradox seems like a philosophical puzzle with no practical implications. But when we consider more involved examples, we may find that constant functions are of little value. Wars, stock market crashes, and similar events are too complex to be modeled as repeating themselves in the same way. Moreover, they tend to affect the probabilities of their own future occurrences. It is in these problems that we have to seriously consider whether "green whenever tested" is indeed more plausible a prediction than "grue whenever tested".

5 References

Goodman, N. (1955). *Fact, Fiction, and Forecast.* Cambridge, MA: Harvard University Press.

Hanson, N. R. (1958), *Patterns of Discovery.* Cambridge, England: Cambridge University Press.

Hume, D. (1748), *Enquiry into the Human Understanding.* Oxford, Clarendon Press.

Kolmogorov, A. N., (1963), "On Tables of Random Numbers", *Sankhya Ser. A*, 369-376.

———(1965), "Three Approaches to the Quantitative Definition of Information", *Probability and Information Transmission*, **1**, 4-7.

Quine, W. V. (1969a), "Epistemology Naturalized", in *Ontological Relativity and Other Essays.* New York: Columbia University Press.

Sober, E. (1975), *Simplicity.* Oxford: Clarendon Press.

Solomonoff, R. (1964), "A Formal Theory of Inductive Inference I, II", *Information Control*, **7**, 1-22, 224-254.

Wittgenstein, L. (1922), *Tractatus Logico Philosophicus.* London, Routledge and Kegan Paul.