NEUROPSYCHOLOGICAL FOUNDATIONS OF PHILOSOPHY

PATRICK SUPPES
Stanford University

1 INTRODUCTION

It is over 200 years since the publication of Kant's *Critique of Pure Reason*. In spite of the many things that Kant understood, and above all, the problems that he saw into so deeply, it is generally recognized that, in any literal sense, the Kantian program is a dead one. There is not going to be, at any time in the future, a serious argument that there is a proper *a priori* synthetic foundation of science or even mathematics. The slow but steady accretion of the case for an empirical view of all human phenomena calls for a revision of much thinking in philosophy that still retains unfortunate remnants needing the kind of critique that Kant gave earlier, but now applied to a wider circle of philosophical ideas. The purpose of this lecture is not to make a systematic analysis of principles of a completely general kind, but rather to give four extended examples of problems that have often been thought of in philosophy, or in mathematics, as not being really empirical in nature. Here they will be presented as naturally so from a psychological and a neural standpoint. An approach to these problems that is purely rational or *a priori* seems, by today's standards of knowledge, mistaken.

It may be thought that what I am advancing as neuropsychological foundations of philosophy is something that is radical and new, but nothing could be further from the truth. In fact, what I have to say in this lecture is very much in the spirit of Aristotle. Let me give just one example to illustrate this point. If one looks for systematic psychological concepts in ancient times, the outstanding example, without any question, is Aristotle's *De Anima*. The treatment of problems of perception and of thinking have no match in any other text of the ancient world. Some credit, of course, must be given to Plato for an early beginning. The fundamental importance of the *De Anima* was fully recognized by many later commentators, including especially the detailed commentary of Aquinas; and the clarifying

paraphrase of Themistius, written in the 4th century CE, widely read even in the Renaissance.

The thesis I am advancing is not meant to be universal in philosophy, but all the same wide ranging. The first example tries to bring out the empirical character of the ordinary use of the concept of truth, and how far the psychological methods by which the truth of ordinary empirical statements is assessed are from the theories of truth we have had in the past from philosophers and logicians.

The second example deals with beliefs, especially the special case of Bayesian priors. There has not been much philosophical discussion of prior beliefs with which I am taking issue, although there are some relevant aspects of epistemology. Instead, I am attacking the absence of deeper psychological considerations on the part of statisticians, economists, and others who believe that a Bayesian approach is a rational way to think about problems of uncertainty and statistical inference. I agree with much of what they have to say. But I find unsatisfactory the thinness of the psychological foundations that are provided, for example, by the forefathers of the modern Bayesian viewpoint, Frank Ramsey, Bruno de Finetti, and Jimmy Savage. They have important and insightful things to say about the foundation of statistics, but the psychological foundations of their Bayesian ideas are left in an undeveloped and primitive state. This is what I address in the second example.

The third example deals with problems of rational choice and rational thinking in general. In spite of having contributed myself in the past to the rational theory of preference, I find the empirical side of the theory weak. The deeper account of how choices are actually made is a matter of extended psychological development of concepts not usually brought to bear in rational-choice theory. What I have to say about choice here applies also, without more detailed consideration in the limited time available on the present occasion, to norms in general. I hold the same kind of empirical thesis about norms that I hold about rational choices.

Finally, in the fourth example, I set forth a psychological thesis about an important aspect of modern mathematics that is troublesome for many people. On the one hand, as part of the foundations of mathematics, there is a very well worked out formal theory of mathematical proof. It is recognized, on the other hand, by working mathematicians that almost all serious proofs in current mathematical research are, for good reasons, not formal proofs. So, the purpose of this example is to stress the psychological nature of verifying – mind you, not discovering, but verifying – the

correctness of informal mathematical proofs. Here again I find much space for psychological thinking and analysis. Even if one might forecast that in the more distant future, an increasing part of mathematics will be checked in a formal way, informal methods will not disappear.

As yet I have said nothing about the brain. I leave the topic of neural phenomena, in particular neural computations, to the end, and will say no more at this point.

2 FOUR EXAMPLES

2.1. Computation of truth (Suppes and Béziau, 2003)

Philosophers discuss at length various theories of truth – coherence theory, correspondence theory, problem of direct reference, sense and denotation, and so on – but, curiously, do not give an account of how we actually perform truth computations, and even less why we are able to perform them so quickly. Philosophers who claim that "Paris is the capital of France" is true because Paris is the capital of France are generally not interested in explaining how we actually compute the answer. But, since such sentences are almost never remembered, or even previously encountered, a computation is necessary.

Logicians also do not solve these problems. If we want to describe how one answers a question like "Is 49+13 equal to 61?", it is certainly wrong to look at the logical foundation of arithmetic, whether it is proof-theoretical or model-theoretical. We answer such a question by using a series of small computational algorithms and tricks, not by looking for a formal proof from a set of axioms or by finding a model in which the axioms are true and 49+13 = 61 is false. In the case of a question like "Is Rome the capital of France?", it is even more doubtful that we are trying to deduce the truth or falsity of the sentence from a set of axioms, or by using a truth-table.

From my point of view it is misleading to say that we are making a *deduction* to arrive at the conclusion that "Rome is the capital of France" is false, unless we emphasize that deduction does not reduce to the narrow meaning of deduction in formal logic. To avoid misunderstanding, it is better to say that we are here trying to describe how we *compute* the truth or falsity of such a sentence.

In a recent book on computational semantics, the authors say:

> The book is devoted to introducing techniques for tackling the following two questions:
>
> 1. How can we automate the process of associating semantic representations with expressions of natural language?
>
> 2. How can we use logical representations of natural language expressions to automate the process of drawing inferences?
>
> (Blackburn and Bos, 2005, p.iii)

Their idea is to find some algorithms to translate natural language into the language of first-order logic to represent the meaning of natural-language sentences, and then to find some additional algorithms to make inferences with these first-order translations. The two steps seem wrong for our purpose. It is highly doubtful that our brains use first-order logic to compute empirical truths. Both AI researchers and computational linguists have been overly-influenced by formal logic. They do not deal directly with the problem of finding the obvious truth or falsity of atomic statements like "Rome is the capital of France".

The theory of the computation of such truths, i.e., the truths of ordinary empirical statements, is an important aspect of how the mind works. It is by no means anything like the whole story of the computations in which the mind, or the brain, is involved. Very much more is required in even the simplest computations of perception. Just think of the necessary computations to decide, perhaps incorrectly, that the image from a certain perspective of a person 200 meters away is indeed an image of your mother. What I shall have to say about computation will be much simpler than that of organizing such perceptual input to form beliefs about what I am seeing. Put another way, because of the great importance of perception in all our activities, we are continually forced to make a dazzling array of computations in processing stimulus input that actually reaches the cortex as electromagnetic signals, reflecting a marked degree of abstraction from the vivid language of ordinary talk about processes and things. It is mind-boggling when first thought about. Indeed, the electromagnetic signals that the cortex processes seem inherently more difficult to understand than the sensible forms of Aristotle's theory, as set forth in the *De Anima*.

In spite of my references to electromagnetic signals, discussion here will be at a still more abstract psychological level most of the time, but I will turn back on several occasions to the brain rather than the mind,

because, historically, the literature on the mind is almost entirely absent any serious theory of computation. So, to continue this general point about computation and how the mind works, all the neurophysiological processes of perception that reduce observed features of things and processes to electromagnetic signals sent to the cortex are ignored in ordinary or philosophical talk about experience. And so modern philosophy of mind is not concerned with the details of how we learn about the ways in which phenomena in the world are connected.

I chose the last word deliberately. The approach to computation about such things and processes, characteristic of our minds, was well recognized by Hume, the godfather of the central mechanism of association, already foreshadowed by Aristotle. What I shall insist on here is the nearly universal role of association as the main method of computation in the brain at the system, but not cellular, level (and in the mind, if you will) in dealing with ordinary experience. The point is an important one, even if there is not space here to muster all the arguments that I think are relevant.

I will make the following general claim about computation. It is sometimes felt that a very clear criticism of behaviorism and connectionism in relation to much behavior is easily made. Rules play too central a role, it is held, to believe an associationist account could be correct. This, however, rests upon a deep mathematical misunderstanding of what can be done with quite simple methods of association or conditioning. It is evident from the many proofs that any computable function (and thus any rule) can be computed by a universal Turing machine, by a universal register machine, or by any of six or seven other devices. Very elementary primitive ideas are quite sufficient, once there is any method of recursion available, to prove that the basic device to do the computing can be quite simple in conception.

All of these remarks are a kind of prolegomena to what I have to say about the computation of the truth of ordinary empirical statements. One point I want to make is that I shall not, in this discussion of truth, distinguish between belief and truth. It is possible to be too zealous, from a philosophical standpoint, and not accept a discussion of the truth of ordinary statements, as opposed to 'Which ones do you believe to be true?' In fact, in much ordinary discourse, claims about belief are mainly used to express doubts about truth, not as a separate point of positive emphasis, for example, "Do you really believe what he said is true?"

Associative networks. Anyway, I want to give a sketch of a theory of how such ordinary computations of truth are made. In doing so, I draw on a

recent article of mine with Jean-Yves Béziau (2003). The basic idea is that the computations are made by an associative network with brain representations of words being the nodes and the links between being the associations. More generally, auditory, visual, and other kinds of brain images can also be nodes.

In the initial state, not all nodes are linked, and there are, in this simple formulation, just two states, *quiescent* and *active*. No learning or forgetting is considered. It is assumed, without being formulated here, that, after a given utterance is responded to as being either true or false, all the activated states return to quiescent. The axioms, which are not stated here, are formulated just for the evaluation of a single sentence, not for giving an account of how the process works over a longer stretch of discourse. The way to think about the networks introduced is that a person is asked to say whether a sentence about familiar phenomena is true or false. It is very natural to ask, and not to have a quibble about 'Do you believe this, even though you don't know whether it is true?' I take examples that are so obvious everyone accepts them as true or false, as the case may be. Simple geography sentences are often used in experiments on these matters. Here is one: *Warsaw is not the capital of Austria*. This sentence input comes from outside the associative network in the brain. I will consider only spoken words forming a sentence, although what is said also applies to visual presentation, as well. So, as the sentence is spoken, the sound pressure image of each word that comes to the ear is drastically transformed by a sequence of auditory computations leading to the auditory nerve fibers which send electromagnetic signals to the cortex. Such signals are examples of those mentioned earlier. In previous work, I have been much concerned with seeing if we can identify such brain signals as brain representations of words. Some references are Suppes, Lu, and Han (1997) and Suppes, Han, Epelboim, and Lu (1999a, 1999b).

The brain activates quiescent states by using the energy for this activation from that brought into the cortex by the brain representation of the verbal stimulus input. With the activation of the brain representation of words by external stimuli, the associations, i.e., links, between activated brain representations are also activated.

Moreover, it is assumed in the theory that energy can be passed along from one associated node to another by a phenomenon characterized some decades ago in psychological research as *spreading activation* (a good reference is Collins and Loftus, 1975). For example, in a sentence about a city like Rome or Paris, some familiar properties are closely asso-

ciated with these cities and the brain representation of these properties may well be activated shortly after the activation of the brain representations of the words *Rome* or *Paris*, even though the names of these properties, or verbal descriptions of them, did not occur in any current utterance. This is what goes under the heading of *spreading activation*. Some form of it is essential to activate the nodes and links needed in judging truth, for, often, we must depend upon a search for properties, which means, in terms of processing, a search for brain representations of properties, to settle a question of truth or falsity. A good instance of this, to be seen in the one example considered here, is the 1–1 property, characteristic of such a word as *capital: x is capital of y*, where here, *x* is ordinarily a city and *y* a country. There are some exceptions to this being 1–1, but they are quite rare and, in ordinary discourse, the 1–1 property is automatically assumed. But this is only one of many other examples, easily given, that arise in ordinary conversation.

One other notion introduced is the notion of the *associative core* of a sentence, in our notation, $c(S)$ of a sentence S. For example, in the kinds of geography sentences given in the experiments referenced above, where similar syntactic forms are given and the sentences are given about every four seconds, people apparently quickly learn to focus only on the key reference words, which vary in an otherwise fixed sentential context, or occur in a small number of such contexts. So, for example, the associative core of the sentence *Berlin is the capital of Germany* is a strongly linked core of three nodes, the brain representations of the three words *Berlin*, *capital* and *Germany*. For such a core I use the notation BERLIN/CAPITAL/GERMANY, with, obviously, the words in caps being used to denote the brain representations, i.e., the three nodes in the associative network. A more complicated concept is obviously needed for more general use.

In the initial state of the network associations are all quiescent, e.g., PARIS ~ CAPITAL, and after activation we use the notation PARIS ≈ CAPITAL. In the example itself we show only the activated associations and the activated nodes of the network. The steps of the associative computation are numbered temporally t_1, etc.

Example: *Rome is the capital of France.*

t_1.	ROME, CAPITAL, FRANCE	Activation
t_2.	PARIS, 1–1 Property	Spreading activation
t_3.	ROME ≈ CAPITAL, CAPITAL ≈ 1–1 Property	Activation
	CAPITAL ≈ FRANCE, PARIS ≈ CAPITAL	
	PARIS ≈ FRANCE	
t_4.	ITALY	Spreading activation
t_5.	PARIS / CAPITAL / FRANCE	Activation
	ROME / CAPITAL / ITALY	
t_6.	TRUE ≈ PARIS / CAPITAL / FRANCE	Spreading activation
	TRUE ≈ ROME / CAPITAL / ITALY	
t_7.	FALSE ≈ ROME / CAPITAL / FRANCE	Spreading activation

This sketch of an example, without stating the axioms and providing other technical details, is meant only to provide a limited intuitive sense of how the theory can be developed for simple empirical sentences. Most important, there is here no account of learning associations. Only an idealized performance setup is used.

But my point should be clear: such computations, or something like them, dominate ordinary discourse, and standard philosophical theories of truth are of little help in thinking about them. Detailed psychological theories of association are more useful.

2.2. Where do Bayesian priors come from? (Suppes, 2007)

Bayesian prior probabilities have had an important place in the theoretical and practical consideration of probabilistic and statistical methods since at least the middle of the twentieth century. In spite of this widespread interest in theoretically using Bayesian priors, and often empirically eliciting them, the analysis of where these priors come from and how are they formed has received little attention.

The absence of such consideration can be seen in the rather laconic views about prior probabilities themselves, as expressed by the three most important foundational thinkers on the Bayesian viewpoint in the twentieth

century, namely, Ramsey, De Finetti and Savage. Quotations are omitted here, but supplied in the article referenced.

What is remarkable about the views of these three foundational thinkers is that none of them ventures very deeply into the psychological or common sense side of how, in fact, subjective probabilities are formed. If we are planning a scientific experiment and ask individuals for their prior beliefs about the design or outcome, almost everyone will agree to this summary of the situation. "Well, if this is an experiment in physics, it's likely that an experienced experimental physicist is going to have a much more interesting prior about the outcome of the experiment than will, for example, the most distinguished professor of philosophy or of English literature." Why is this so? Because we believe, well beyond any requirements of coherence or consistency, as it is sometimes called, there is the really much more important matter of the background experience which led to the formation of a given individual's prior.

It is reasonable to accept the lack of detailed psychological theory of the mechanisms by which prior probabilities are formed, but it is less excusable that there is an almost total absence of a detailed discussion of the highly differentiating nature of past experience in forming a prior applicable to a new experiment or, more generally, almost any action about to be undertaken. Note, of course, that the use of the term 'prior' is, in some sense, misleading. Of course, the partial beliefs we are eliciting in the case of an experiment or an action are prior to the experiment being conducted or the action being taken. But, they are not prior to experience relevant to the experiment or action. There is, in fact, usually much relevant prior experience. There is, if you wish, a beginning to the experience; we might say that we want to go back to the very beginning and consider only priors while still in some perfect state of ignorance. This is not a point worth quarreling about, but it is a point worth noting in terms of our linguistic usage. In fact, in talking about priors, we almost always accept that there has been experience prior to the elicitation of the prior. I will not return to this point, but I do think its consideration is badly missing in the usual discussion of priors. (This is an old point of mine about the obviously unsatisfactory character of many possible priors, Suppes, 1956, p.72.)

It is often said priors express, at least partly, differences in taste. We can agree that there will be differences in taste, even among experts. But it is also essential to make the point that we can properly and empirically assess whether or not the accuracy of priors of experts in a subject are better or worse than beginners and, in fact, who among the experts has a better

record. Again, although this is a worthwhile topic, it is one for another occasion.

What I want to focus on here is what kind of account, even if necessarily schematic, can be given about the psychological mechanisms back of the formation of our Bayesian priors.

Nearly as common as our empirical statements, which are implicitly enunciated in such a way as to make clear that they are held, without any question, to be true, is our more tentative estimation of future events, or statements about past events, whose occurrence is uncertain in our minds. A familiar idiom of great importance, both in ordinary conversation and in the theory of probability, is the rich notion of expectation. So we speak of expecting to be home in 30 minutes, expecting to lose five pounds in the next two weeks on our current diet regime, or expecting to purchase a new car for less than we had anticipated, because of the highly competitive nature of the market. In all these cases, expectation is our way of dealing with events that are measured in terms of quantity. So, the outcome is not just a yes-or-no occurrence, but something to which is attached a quantitative measure – time measure, weight measure, money measure, etc. I will say more about expectations later.

At the moment, I will extend the analysis of the previous section by just considering events that occur or do not occur. Moreover, to simplify the formulation and discussion of examples, I restrict myself to events and their brain representations rather than consider sentences. This change is made for simplification, but also matches ordinary practice, which uses an event-formulation more than a statement-formulation as the basis for probability claims. This matches theoretical developments as well. The statement-formulation of probability, as in confirmation theory, developed mainly by philosophers, is not widely used, for many reasons, in systematic theoretical and scientific applications of probability theory. The set-theoretical framework of events is more common and useful for many familiar reasons, which I will not repeat here. (The foundations of this set-theoretical view, written in a way that is meant to be accessible to philosophers, is given in Chapter 5 of Suppes (2002).)

A second point about what I shall limit myself to doing here is that I will not generalize in complete form the axioms, referred to in the previous section dealing with the case of the truth of simple empirical statements, but only sketch their formal developments, which otherwise would require a much larger excursion into technical details than is appropriate. So, the remainder of the discussion will be at an informal level. I do give, in the

article referred to (Suppes, 2007), a detailed example that represents a relatively simple restricted set of theoretical ideas that have also been extensively tested experimentally.

So, as we turn from truth to the estimation of probabilities, especially Bayesian priors, there are a number of observations with which I want to begin. The first is that such priors are based on a variety of experience, not on the sharp outcomes of well-planned experiments. The fundamental point is that we come to the design of experiments with such Bayesian priors well developed. It is they which guide, in many different ways, our thinking about the design of such experiments. It is, of course, a fundamental point about the design of experiments that we do not come to them with well-prepared algorithms, mechanically applicable, to hand us the design on a silver platter. It is not at all that way. Designing an experiment and executing that design are rather like designing a house and then building it. There are many practical decisions that must be taken along the way that are no part of any set of known algorithms and that are inescapable in actual work. Our prior knowledge and experience are the most helpful things we have. I emphasize experience rather than knowledge, because much of this experience is not consciously articulated – that marks the difference between amateur experimenters and experienced ones. Imagine turning an amateur loose in a modern physics laboratory. In almost any aspect of experiments now conducted in physics, from quantum entanglement to superconductivity, prior experience is the key to success. This kind of experience is gained from the kind of apprenticeship that is very similar to that found in any specialized work in ancient China, Egypt or Mesopotamia. More generally, this kind of prior experience is necessary in every aspect of ordinary affairs requiring some kind of learned competence, from driving a car to cooking a decent meal, or installing and using a photodetector.

Important in this formulation is the recognition that, by *Bayesian priors* we do not mean the beginning of the beginning, but the beginning of the end, in the sense that when we come to any of these tasks, we come with much developed skill and experience. The prior refers to our state of knowledge, skill and practical competence as we face the task at hand. The term *Bayesian prior* applies, of course, particularly to experiments, where we may be anxious to develop a body of knowledge into which the likely outcomes of the experiments will be accepted by all, or almost all, competent persons. This acceptance should lead to a modified posterior probability, which can, in turn, be taken as a future prior.

My theoretical point is that the basic mechanism of forming these priors, as in forming many other things, is that of association. I emphasize again the central principle that the mechanism of association is, if not universal, nearly universal, in the acquisition of knowledge and skill in animals and humans alike. It is not something special that is turned on for the purposes of acquiring a Bayesian prior, but is a deep mechanism of organisms from *Aplysia* to *Homo sapiens*. What is also apparent from what I have just said is that, though we may be hopeful of stating some general laws of association, the detailed analysis of the way associations are built up in everyday heterogeneous experience will not be practical in any detail. Just as in the external world, the testing of physical theories in heterogeneous experience scarcely occurs at all. As everyone knows who has thought about the problem for any length of time, it is impossible to consider writing down the differential or difference equations governing even the detailed motion of the leaves in the tree outside my study as I write these words. So, no apologies are needed for being able to find only partly systematic verification for the claims that are made for association, because this is no different from the claims that can be made for any general physical law. On the other hand, this is not to make the claim that the mechanisms of association are currently as well formulated and as well understood as the mechanisms postulated and tested in many parts of physics.

However we model things in particular, there are two general assumptions we need. The first is that, as we build up the strength of associations, there is an independence-of-path assumption that holds, at least approximately. This means that the exact historical path we followed in the experience of building up an associative connection does not strongly affect the strength of the association. The same strength can be reached by many other paths. The important point is that we do not need to know the detailed history of the past to infer the current intensity. This would, it seems plausible, impose an impossible burden on memory to require that we do indeed continue to keep track of each past increment to the intensities of associations. I think it is fair to say that the evidence is substantial that, in spite of the vast capacity of human and animal memories, brain computations based on such detailed histories would be too cumbersome for practical biological use.

The second approximate general assumption is the fading influence of the past, which means that our system of knowledge, skills and actions is nearly ergodic. In other words, the influence of the past fades away in some exponential fashion. What I have just said about memories would be

an example of this. The accumulated experience of the past is kept in the relative intensities of association, but the memory of each increment, positive or negative, to these associations, is not kept. It is necessary, of course, to say that this ergodic property is approximate, because there are some salient memories that are remembered, but, compared to the total experience, their relative scarcity is evident.

Simple example. So, the mechanisms for estimating probabilities can be exemplified in a familiar example. Someone asks me the question 'Will it rain in Paris tomorrow morning?' I compute my answer by using the associative network of features and patterns of features in memory, activated by the brain representations of the words 'rain', 'Paris', and 'tomorrow morning'. This activation is now postulated to be more complicated than that of the previous section. In particular, it seems essential and fundamental that some concept of intensity be used to represent the varying strengths of associations. We shall do so using notation familiar in current work in neural networks by speaking of varying weights w_i, where w_i is meant to represent an intensity built up from past associations. The outcome, now, of computation in the associative network of the estimated probability can have in summary form, a simple representation:

$$\text{rain: } \sum w_i = r$$
$$\text{no rain: } \sum w_j = \bar{r}$$
$$\text{probability of rain } = \frac{r}{r + \bar{r}}$$

Of course, what I have shown here is the outcome of the computation, not how such a computation might actually be made by the brain.

More on representation and association. We can see from this simple example the main ways in which the assumptions considered in the previous section on truth must be extended and changed. First, there must be a clear mechanism for the build up of the intensity of activation in terms of change. Second, we must, explicitly, not consider just a given event, but, for the estimation of its probability in some form, the probability of the opposite happening. In some cases, there may be schemes for directly estimating only the probability of the event, but it is most natural, when weights are used, to determine probabilities from simple relationships between the weights, as in this example. Another important point of extension of the setup for truth, especially as we move on to the estimation of probability, is that it is necessary to think in terms of brain representations

of other things than words. Words will continue to be important and they seem fundamental in much of our own mental activity. But, certainly, it is wrong to think just in terms of words as we move to nonhuman animals, where the brain representations of past experiences must obviously be non-verbal in form. But I see no problem with the representation in the brain of many different kinds of things – words, past events, past scenes that we have seen or things that we have heard. For example, the way in which passages of music, familiar from the past, generate brain representations that associate to a variety of emotional experiences is well known.

This remark may well raise a question in the minds of some, especially those oriented towards analytical philosophy or cognitive psychology: "But where are the concepts? Why speak of the brain representation of the word *rain* as opposed to the brain representation of the concept rain?" My answer to that is that we do not have direct brain representations of concepts, in the sense that we can physically identify events in the brain and say 'Ah, there is the concept of rain'. The events localized in time and space, so to speak, that we identify will be representations of concrete events, for example, instances of rain, or instances of words. As many readers will recognize, what I am upholding here is the exemplar theory of concepts. This is the theory that originates with Berkeley and Hume and is stated so elegantly in an early passage in Hume's *Treatise of Human Nature* (1739). This theory is also prominent among current theories of concepts. It stands more or less as a co-equal with the prototype theory and the classical definitional theory. (For a current account of psychological research on these three theories, see Murphy, 2002. But the most elegant exposition is given in the early pages of Hume's *Treatise*.) We need not be so sarcastic and sardonic as Berkeley, making fun of Locke's theory of abstract or general ideas, when we affirm there is much in the activity of the brain, as now understood, at least, to confirm Berkeley and Hume and the current exemplar theory. (For some detailed neural data on this point, see Suppes et al., 1999a.) For those who like their concepts clean and simple, there is much that is disturbing about the associative-network view I am proposing. Such networks have a natural inexhaustible complexity. William James puts the matter nicely in his chapter on association in the *Principles of Psychology*.

> The jungle of connections *thought of* can never be formulated simply. Every conceivable connection may be thought of – of coexistence, succession, resemblance, contrast, contradiction, cause and effect, means and end, genus and species, part and whole, substance and property, early and late, large and small,

landlord and tenant, master and servant, – Heaven knows what, for the list is literally inexhaustible.

… If pure thought runs all our trains, why should she run some so fast and some so slow, some through dull flats and some through gorgeous scenery, some to mountain-heights and jewelled mines, others through dismal swamps and darkness? – and run some off the track altogether, and into the wilderness of lunacy? Why do we spend years straining after a certain scientific or practical problem, but all in vain – thought refusing to evoke the solution we desire? And why, some day, walking in the street with our attention miles away from that quest, does the answer saunter into our minds as carelessly as if it had never been called for – suggested, possibly, by the flowers on the bonnet of the lady in front of us, or possibly by nothing that we can discover? If reason can give us relief then, why did she not do so earlier?

(James, 1890/1931, I, pp.551–552)

The patterns of association are rather like the patterns of railroad tracks built up over a century. Many are still used. Others are only barely visible and for some only the right-of-way can now be faintly seen. There is no sharp definition as to what the current state is, depending on use, on property rights, etc. And, so it is with associative networks. The main tracks we know well. They lead to the simple truths exemplified in the previous section. The common tracks for common talk about probabilities and expectations in everyday matters are of lesser strength in association but still quite manifest. The wilder reaches of association are another thing, still there and still important for some purposes, but seldom used. There is no drawing a line of any analytical precision as to where the associative network of a given concept ends and those of new ones begin. The line to be drawn is as arbitrary, if precision is insisted upon, as is that between the analytic and the synthetic.

2.3. Habits as the basis of the theory of rational choice (Suppes, 2003)

Unconscious nature of thinking. Our mental concept of ourselves is above all that of self-aware thinking beings. The pinnacle of rationality is systematic deliberation about ends and means for achieving those ends. From Aristotle to the present, practical reasoning has been a focus of attention in philosophy, but in spite of the acuity of much of what has been written, the complexity and sophistication of the kinds of problems considered as presenting issues for the application of practical reasoning have been limited. What has been especially missing has been attention to the large psychological literature on the nature of thinking, and in particular, the literature

concerned with the thinking processes involved in making serious and seemingly deliberate choices that involve major personal goals.

Contrary to much folklore psychology and the implicit assumptions of many philosophers, we are almost entirely unaware or unconscious of our detailed thinking processes. What we have excellent knowledge of is the results of thinking, often of partial results that constitute major steps in reaching a final decision about an important matter. Here is a relatively brief survey of the many kinds of experimental studies supporting these conclusions. They set scientific psychology in opposition to folklore psychology and numerous philosophical ideas and ideals about the rationality of practical reasoning. In fact, it is important not to imply a serious restriction to practical matters. The proper view of the unconscious nature of thinking processes applies to finding solutions to theoretical problems as well.

Two seminal articles on these matters are that of Nisbett and Wilson (1977), whose title is "Telling more than we can know: verbal reports on mental processes" and Wilson (1985), whose title is "Strangers to ourselves: the origins and accuracy of beliefs about one's own mental states." These articles survey in depth a number of experimental and nonexperimental empirical studies over many years, including their own work. I give a brief summary here.

In the first category I mention studies concerned with the inability of individuals to answer "why" questions. Gaudet (1955) found that respondents could not explain why they liked particular political candidates. Ranging far afield from this, Kornhauser and Lazarsfeld (1955) found that respondents could equally not explain why they liked certain detergents for laundering purposes. Lazarsfeld (1931) found that respondents could not explain why they chose a particular occupation and, in a similar vein, Davis (1964) found respondents could not explain why they chose to go to graduate school. Further back in time, Burt (1925) found respondents could not explain why they became juvenile delinquents or, in terms of more positive decisions, Goode (1956) found respondents could not explain in any reasonable way why they got married or divorced. Rossi (1955) found respondents unable to explain why they moved to a new home.

In discussing these examples on several different occasions, I have chosen to expand upon the example of buying a new house. This is a traumatic and difficult process for nearly everyone who has been involved in it. Almost without exception, explanation of the particular choice made is woefully inadequate. This does not mean that certain constraints do not ob-

tain. Individuals are quite competent to state constraints, such as location from schools, overall cost, age of the house and other such factors contributing in a significant way to the final decision. It is just that no overall rationale for the decision taken is ordinarily given. The usual reason is that most individuals, or families, who are selecting a new home, make a very wide search for candidates. They end up with a smaller list with the property that no one dominates all the rest. Consequently, the final decision is based upon something different from the application of a final, solid constraint or a detailed, explicit computation.

Another class of studies, oriented toward theoretical rather than practical problems, concerns individuals' reports on problem-solving processes. Ghilesin (1952) collected data on creative problem solving, as he put it, from Picasso to Poincaré. He emphasizes that production by a process of purely conscious calculation seems never to occur. A classic study of Maier (1931) on combining extension cords on a ceiling for electrification purposes shows how unconscious problem-solvers usually are of their pursuit of a solution. In mathematics there is widespread recognition that theorem-proving of any difficulty depends upon imaginative leaps very similar to memory retrieval, but clearly computational in character. The key idea, just like that of retrieval of a memory, comes into consciousness with no trace at all of how it was arrived at. There are numerous famous anecdotes by scientists and mathematicians about this process. I shall not review them here, but almost everyone is aware of what Hadamard and Poincaré have claimed in this respect. I have never heard a serious mathematician deny that this important role of unconscious processes was in fact always at work in obtaining any significant mathematical result. Here is a short famous quotation from Hadamard (1945).

> One phenomenon is certain and I can vouch for its absolute certainty: the sudden and immediate appearance of a solution at the moment of sudden awakening. On being very abruptly awakened by an external noise, a solution long searched for appeared to me at once without the slightest instant of reflection on my part – the fact was remarkable enough to have struck me unforgettably – and in a quite different direction from any of those which I had previously tried to follow. (Hadamard, 1945, p.8)

The attempts to explain this lack of awareness have produced a large number of new experiments, hypotheses and theoretical analyses from psychologists. Let me just summarize some of the reasons given for why we are unaware of our unawareness. The first is a confusion of content and

process. This is not a separation usually made in ordinary talk about decision making, why we have chosen a certain goal or adopted certain means for achieving a certain goal. Second, we have detailed private knowledge of ourselves that is obviously not accessible to anyone else. We can confuse this information with the processes of thinking, because these processes are naturally intertwined with the data that are more or less private for every person. Each of us knows private historical facts about his own thought and action that can affect his thinking processes. Moreover, an individual can tell you his focus of attention at any given moment, which is in itself something quite different from an account of his thinking processes, but is natural to confuse with those processes. Still another factor is private, intermittent awareness of various sensations. We can be aware of seeing a car in the distance or a person nearby missing a step. Recording these observed objects or events can be mistaken for the process of thinking about them.

Perhaps most important, almost all of us are capable of describing coarse intermediate steps in complex problem solving. Good examples are the many steps taken in buying a house, from surveying various neighborhoods, calling an agent, making an escrow deposit, closing the bank loan, to the final dramatic act of moving in. These intermediate steps are intermediate results, easily externally described, but not so for the associated thinking processes. Moreover, such results are at the same time easily confused with the processes themselves, because we do not naturally separate our successive processes of thinking from our successive intermediate results.

What I want to emphasize is this. A theory of rationality that is posited on some exemplary style of rational deliberation, conscious, measured and complete, is utterly mistaken as a psychological account of how any of us go about making decisions about practical problems or solving theoretical ones.

Fantasies of expected utility computations. It is not just the philosophers of practical reasoning that have been mistaken, but it is also the economists and statisticians who have bought into the image of endless rational computations. The further the reach of the computations, the greater the sin of psychological omission in formulating the theoretical ideas. Perhaps the most excessive brand of this is Savage's (1954, p.14) famous fantasy of utility functions over possible states of the world, and the related and intertwined fantasy of de Finetti (1937/1964, p.146) that once we have a probability distribution, all future revisions of thought processes will be by

conditioning only, that is, strictly in the sense of probability theory. Of course, they were both too smart to hold that these fantasies could be realized.

For reasons too numerous to enumerate here, the number of actual long-run calculations ever made is negligible. Keynes had it right. The important fact about the long run is that in the long run we are all dead.

The actual computations we do are fragmentary, occasional, contextual, driven by associations internal and external. A much better guide to thought than the utilitarian principle of maximization taken in its raw form is William James's account of the stream of thought in chapter IX of his *Principles of Psychology* (1890/1931). Here is one passage.

> Now we are seeing, now hearing; now reasoning, now willing; now recollecting, now expecting; now loving, now hating; and in a hundred other ways we know our minds to be alternately engaged. But all these are complex states. (James, 1890/1931, I, p.230)

As I will argue shortly, our computations are built up from myriads of associations, intertwined with our past in ways that we can no more understand in detail now than we can explain how we retrieve a familiar name or a well-known fact from memory. It is why I like to say that when it comes to human computations, fragmentary and associative in character, Proust is a better guide than Turing. Here is a quotation that illustrates this well, from *Time Regained: In Search of Lost Time*, the last part of Proust's extraordinary novel (1927/1999).

> All day long, in that slightly too countrified house which seemed no more than a place for a rest between walks or during a sudden downpour, one of those houses in which all the sitting-rooms look like arbours and, on the wall-paper in the bedrooms, here the roses from the garden, there the birds from the trees outside join you and keep you company, isolated from the world – for it was old wall-paper on which every rose was so distinct that, had it been alive, you could have picked it, every bird you could have put in a cage and tamed, quite different from those grandiose bedroom decorations of today where, on a silver background, all the apple-trees of Normandy display their outlines in the Japanese style to hallucinate the hours you spend in bed – all day long I remained in my room which looked over the fine greenery of the park and the lilacs at the entrance, over the green leaves of the tall trees by the edge of the lake, sparkling in the sun, and the forest of Méséglise. Yet I looked at all this with pleasure only because I said to myself: "How nice to be able to see so much greenery from my bedroom window," until the moment when, in the vast verdant picture, I recognised, painted in a contrasting dark blue simply because it was further

away, the steeple of Combray church. Not a representation of the steeple, but the steeple itself, which, putting in visible form a distance of miles and of years, had come, intruding its discordant tone into the midst of the luminous verdure – a tone so colourless that it seemed little more than a preliminary sketch – and engraved itself upon my windowpane. And if I left my room for a moment, I saw at the end of the corridor, in a little sitting-room which faced in another direction, what seemed to be a band of scarlet – for this room was hung with a plain silk, but a red one, ready to burst into flames if a ray of sun fell upon it. (Proust, 1927/1999, pp.9–10)

This long passage from Proust shows why he is a better guide to human computation than Turing. The true complexity of much, if not most, human computing is to be found in perception. The human visual system may be the most complicated system in the universe, after the brain itself. And our continual attention to vision, seen from an unusual angle, in Proust's highly particular perceptions and associations, is characteristic of much of our waking hours, even if we do not usually focus on what we see as intently as in Proust's account. This primacy of perception is testimony to the relative ease of building digital computers compared to the great difficulty of constructing artificial visual systems. The gap between the richness and complexity of perception and thought, so well described by James and Proust, compared to the crude oversimplifications characteristic of any attempt at direct expected utility computations over possible states of the world is an important source of skepticism about the latter.

Habits. There is a scent of *tabula rasa* about the approach to rational choice via maximizing expected utility. It is as if the organism has a simple, uncomplicated structure, whose behavior can be maximized in the way that a simple physics problem can be solved by finding a maximum or minimum of an appropriate quantity. For biological organisms, beginning even with the simplest, nothing could be further from a sensible way of thinking about their behavior. The complexities that can be invoked at this point are much too numerous to be pursued in any detail, but there is one class of phenomena that may be seen not only in mankind but in animals up and down the hierarchy of evolution or complexity. These are the effects of learning on the long-term behavior of an animal. There is, however, a better term, older, and also very much a part of folklore psychology, although not well developed. This is the concept of a *habit*. Some things that we call habits are undoubtedly purely instinctual, that is, are unlearned and encoded in the genes somewhere in the DNA. Most things, however, that we call habits represent an interaction between the genetic structure of

156

an animal and the environment in which it develops and continues to exist. Habits are superb examples of learning, but I want to put the emphasis here on the results of learning, rather than on the learning itself.

Before I say more about habits, let me put my cards face up on the table, so that it will be clear how I am using the concept of habits to help characterize rationality. Habits constitute restraints, in the standard mathematical sense of constraints, on the choices we make. We do not consciously think of our habits in making choices, but concentrate, so far as we exercise conscious discrimination at all, in choosing one thing rather than another, in such a way as to satisfy the appropriate constraints. For example, I am at the stage of my life where I very much prefer wine to beer. At an ordinary dinner in a restaurant, faced with a menu, I only think about the choice of wine, and almost never consider beer. I do not go through any deliberate, rational analysis of the virtues of wine over beer, because of the constraint already established by long-settled habits. I accept the constraint without even thinking or being conscious of it. I can, of course, at another time and for another purpose, make myself conscious of having this constraint. But the important point is that in the act of choosing itself, we do not ordinarily pay conscious attention to the habits we have.

This is not to say that such conscious occasions can never occur. It is the stuff of family drama and the essence of many good novels for a person, real or fictitious, to face up to habits that must be broken, in order to make a choice that is much more important and meaningful to the person than any casual breaking of habits of old. But this is the exceptional situation – one that we can, of course, describe. Yet it is important to get the usual regime of choosing properly thought out. In fact, in the context of this article, I will not attempt to give a serious discussion of when we want to breach our constraints, that is, our habits, and go for something unusual, challenging or even frightening. This is an important topic, but one that can be left to the side, because of the low frequency of such choices, and the necessity of having a much better view of the usual kind of choices we make, from the dramatic ones of buying houses to the trivial ones of choosing glasses of wine.

So, I emphasize, the habits of a lifetime, as the saying goes, present constraints that are ordinarily satisfied. But the constraints do not fix the choice. My strong constraint of always choosing wine, and never beer, does not in any way determine the particular choice of wine on a given occasion.

You may think that I am next going to say that we have come upon the proper role for maximization, namely, to maximize our choices subject to the constraints of habits. But I will not even accept the traditional theory of maximizing expected utility in this reduced role. To anticipate what I will say later, and to give you a sense of the organization of the ideas about rationality I am presenting, the next step after habits is to let the associations of the moment make the choice as freely and as easily as possible. I will not say more about these associations yet, but this is a prelude to what is to replace, not just maximization, but even satisficing, the central concept of Simon's (1955) well-known theory of bounded rationality.

Now back to habits. Much of what I want to say in the context of the present article about habits is said better and in more detail in chapter IV of James's *Principles of Psychology*. I shall not attempt a faithful summary of his ideas, but only emphasize points that are relevant to the characterization of rationality, and I do not claim that what I say is anything like a faithful paraphrase of his thoughts.

The first point is that habits are really physical and already present in nonanimate matter. What we ordinarily think of as certain material properties correspond to what we would call habits in animals. But particle or animal, the habit should be thought of as something physically embodied in the nervous system, and in the muscles, where appropriate. The only real difference on this score between animals and inanimate objects is the much greater mutability of habits in animals. James has a wonderful quote from someone else about the many ways in which matter itself is not immutable. The examples are particularly from designed objects, which have a special property. This is the second point: such objects function better the more they are used. Engines, locks, hinges on doors and the like improve with age, up to a point of course. Let me quote James (1890/1931, p.112), "habit simplifies the movements required to achieve a given result, makes them more accurate and diminishes fatigue." The ironic thing about this aspect of habit is to recognize the importance of efficiency and yet to realize how little it is ever given its pride of place in the discussion of such matters by utilitarians. Habits, indeed, are themselves utilitarian in the deepest sense of that word, namely, in their clearly useful contribution to doing things.

The third property to be mentioned, one of importance in connection with mistaken notions of rational deliberation, is that habits diminish the conscious attention with which acts are performed. In more domains of experience than can be named, only the inept, the awkward and the untrained are conscious of their performances. The accomplished, the gifted and the

well trained are not. And so it is with choices. The final process of choosing is one that is properly left unconscious, once the first round of constraints that are either habitual, or deliberately modified for application to a new situation, have been satisfied. The final reduced choice set should be one worthy of unconscious contemplation and free association. Now many will think that my phrase 'unconscious contemplation' is really overdoing it. Only the mindless choose this way. The data show otherwise. Only the inept are mindful of their final choices, to put the matter in the most controversial way, but one about which I am all the same utterly serious.

Finally, I cannot forego one more quotation from James about the important social role of habits. This topic lies somewhat outside my main focus here, which is on individuals, but a theory of rationality that ignores the social framework, of one kind or another, in which all of us live, is a Robinson-Crusoe view that is clearly a reductive absurdity.

> Habit is thus the enormous fly-wheel of society, its most precious conservative agent. It alone is what keeps us all within the bounds of ordinance, and saves the children of fortune from the envious uprisings of the poor. It alone prevents the hardest and most repulsive walks of life from being deserted by those brought up to tread therein. It keeps the fisherman and the deck-hand at sea through the winter; it holds the miner in his darkness, and nails the countryman to his log-cabin and his lonely farm through all the months of snow; it protects us from invasion by the natives of the desert and the frozen zone. It dooms us all to fight out the battle of life upon the lines of our nurture or our early choice, and to make the best of a pursuit that disagrees, because there is no other for which we are fitted, and it is too late to begin again. (James, 1890/1931, p.121)

We don't have to accept or use all of James's examples. We can easily write new ones, suitable for our own age and technology, but his point is understandable without any changes needed.

Entropy and free associations. A habit that is deterministic will, of course, have an entropy rate of zero. In my familiar example of usually choosing wine over beer in a restaurant, the entropy rate of my responses, at the concrete level of the kind of wine, vintage and winemaker selected, will not be zero. Notice that the level of abstraction selected will vary the entropy rate. It is also part of my philosophy of these matters that there is no ultimate concrete specification, so that any level selected reflects some kind of abstraction. As we eliminate vintage, say, first, then winemaker, and then kind of wine, we expect the entropy rate to decrease, so that finally, if we have only the choice of beer, wine or soft drink, as the three

possible choices, my entropy rate is close to zero. (An interesting question for consumer-behavior studies is what level of abstraction is of the most interest in calculating entropy rate.)

There is a deeper question and one I am not yet entirely clear about, but fundamental to the ideas I am working on. This is what is the proper level of abstraction, in terms of what is represented mentally (or in the brain). So, after making some determination of habit, if the entropy rate is not zero, room is left for free associations. It is especially the free associations that we expect to be malleable and therefore subject to transient changes in stimulation; such expectations are also characteristic of firms that vie for shelf space to advertise their products.

My tentative answer to the level of abstraction of the associations is that it just depends on the strength of resemblance or similarity between the mental (or brain) images, on the one hand, and the stimuli on the other. And, in fact, it is a mistake of mine to introduce the misleading idea of abstraction. It is better to introduce different relations of similarity, which we can use to make corresponding, but more psychologically realistic, claims. In other words, any use of abstraction should be backed up by a working concept of similarity or isomorphism to define the particular level of abstraction.

Both the concepts of habit and of free association can be applied with varying definitions of similarity or isomorphism. Note that the two, habit and free association, must go together, if we want to complete the study of choice. For example, to use again my familiar example, if we consider just my standard choice of wine over beer, habit completely accounts for my choice at this level and there is no room left for free association. But if we make the isomorphic or similarity relation more detailed, there is. In fact, in the present formulation of ideas, whenever the level of characterization of a habit has nonzero entropy, the remaining nontrivial choice set leaves room for free associations.

This remark leads to the natural question of how to distinguish between habits and free associations. Can we just define a relation of isomorphism or similarity at any level and thereby mark a distinction, so that we distinguish only relative to such a relation? In some ways this seems a good choice, for after all, according to the ideas being advanced here, association or the special case of conditioning, is also at the basis of habit, except possibly for some small part that is genetic in character. Is such a complete relativization of the distinction between habit and free association a satisfactory answer? I do not think so. For, it seems to me, it is important

also to separate the ephemeral quality of free associations from the lasting quality of habits. This separation can be made by introducing further distinctions among the similarity relations used, based on their temporal character. I do not pursue the formal details here.

Associations as natural computations. From a philosophical standpoint, the great opposition to the fundamental mechanisms of the mind being just associative computation and memory is the Kantian line of transcendental idealism grounded in the *a priori* synthetic. But it is important to note that Kant thought that Hume was right in what he claimed empirically for association (*Critique of Pure Reason*, 1781/1997, A100). It is just that he did not accept that Humean empiricism was ultimately enough as a foundation for science, especially for Newtonian mechanics and mathematics.

Writing a hundred years later, William James is an enthusiastic critic of Kant's grounding of science with necessary *a priori* synthetic principles. Here is a passage expressing his thought well.

> … The eternal verities which the very structure of our mind lays hold of do not necessarily themselves lay hold on extra-mental being, nor have they as Kant pretended later a legislating character even for all possible experience. They are primarily interesting only as subjective facts. They stand waiting in the mind, forming a beautiful ideal network; and the most we can say is that we *hope* to discover outer realities over which the network may be flung so that ideal and real may coincide. (James, 1890/1931, I, pp.664–665)

The passage comes nearly at the end of the 2nd volume of James's deep and majestic survey of 19th-century psychology. But it is not an isolated few lines. James attacks again and again Kant's transcendental idealism and his attempted *a priori* grounding of knowledge.

Moving ahead to more recent developments in psychology, the special case of association that is important, in the first half of the twentieth century in the development of psychology, is, of course, conditioning. The concept of conditioning dominated thinking about almost all aspects of psychology from the first decade of the twentieth century to the second half of the century. It ended only with the linguistic revolution of Chomsky and others, and the subsequent development of a cognitive psychology that, to a large extent, has emphasized the role of rules over associations as the basis for thought. This regime, which was prominent from about 1965 to 1980, has had, as its hallmark, the replacement of nonsymbolic by symbolic thought. The decline of this line of theory began around 1980 with

the introduction of nonsymbolic computational processes, so characteristic of modern neural networks. More than two centuries after the death of David Hume in 1776 we again find ourselves returning to associations, now often in the form of neural networks. Currently they occupy the dominant place in the conception of the mechanisms of thought. Not everyone will agree with the formulation I have just given. Many will claim that it is still just too strong to say this, that there are other modes of thinking that remain of great importance. I am skeptical of that. I am happy to push the thesis that those other modes are themselves splendid examples of conditioning, for example, the mental computations of arithmetic, the algorithmic rules we all learn early. If we turn from such algorithms with the contempt with which many cognitive scientists and some mathematicians do, then the response is even better. Surely the evidence is that the best and hardest mathematical proofs arise, not from some linear, nicely formulated line of explicit reasons, but from random, scattered, jumbled associations of the kind mentioned in the passage from James and the one from Hadamard. Only later is an orderly exposition of justification found.

To push these ideas further, in 1969 I gave a clear mathematical proof that, just from ideas of stimulus and response, we could generate finite automata (Suppes, 1969). In a later article (Suppes, 1977), I showed how to extend these ideas to an arbitrary Turing machine, all operating by conditioning, that is, by special cases of association. The argument is amplified in Suppes (2002, ch.8). From a psychological standpoint, these constructions of finite automata or simulated Turing machines are too simple. No doubt the actual computational processes in the brain using associations extensively are more devious and complicated. Moreover, we do not begin language learning with a mind that is a *tabula rasa*. Much structure and related processing is constrained by our common genetic inheritance. It is then above all association or conditioning that shapes the further development.

There is one additional point I want to make to those who remain skeptical about association. Think about your own methods of memory retrieval, and then try to give a theory that does not deeply involve processes of association.

Freedom of association. As some may note, the title of this section is meant as a double entendre. On the one hand, I have in mind associations in the brain, and on the other, the great historic libertarian demand of freedom of association for the individual. But it is the brain about which I am serious at this point. Let me be explicit about what I want to mean by *free-*

dom of association. I have in mind a hierarchical conception of how we make rational choices. To begin with, we must satisfy our habits. With satisfaction of the constraints given by habits or some specific computations, we are then left with an unresolved set of choices. How should we choose from this set? The classical utilitarian method is by maximizing utility. The classic algebraic theory I consider a hopeless enterprise, for reasons already given. The rational individual, who satisfies the constraints of habit, is one who is freely associating and choosing that one of the remaining available set of options that seems most attractive, based on past associations that are brought up, as can be the case in buying a house, or, in other instances, by the association to anticipated events. Often, a glimpse at something attractive nearby sets off the train of associations. Belief in the relatively high frequency of this last case is a fundamental tenet of advertising.

The immediate reaction of some readers may be to challenge this probabilistic mechanism of choice as normal. They may recall (perhaps I should say, *associate*) their earlier encounter with the literature of psychoanalysis and its emphasis on the central role of free association in interpreting dreams or analyzing repressions, slips of the tongue and many other phenomena. But the central role of association in our mental life was not a Freudian discovery. It goes back at least to Aristotle. Here is Freud describing the associations arising from the interpretation of a dream:

> And next, we obtain these associations. What they bring us is of the most various kinds: memories from the day before, the 'dream-day', and from times long past, reflections, discussions, with arguments for and against, confessions and enquiries. Some of them the patient pours out; when he comes to others he is held up for a time. Most of them show a clear connection to some element of the dream; no wonder, since those elements were their starting-point. (Freud, 1971, p.11)

It does not sound much different from one of the earliest references to associations in various passages of Aristotle's *On Memory and Recollection*. For example,

> It often happens that one cannot recollect at the moment, but can do so by searching, and finds what he wants. This occurs by his initiating many impulses, until at last he initiates one such that it will lead to the object of his search. For remembering consists in the potential existence in the mind of the effective stimulus; and this, as has been said, in such a way that the subject is stimulated from himself, and from the stimuli which he contains within him.

But one must secure a starting-point. This is why some people seem, in recollecting, to proceed from *loci*. The reason for this is that they pass rapidly from one step to the next; for instance from milk to white, from white to air, from air to damp; from which one remembers autumn, if this is the season that he is trying to recall. …

If one is not moving along an old path, one's movement tends towards the more customary; for custom now takes the place of nature. Hence we remember quickly things which are often in our thoughts; for as in nature one thing follows another, so also in the actualization of these stimuli; and the frequency has the effect of nature. …

That the experience is in some sense physical, and that recollection is the search for a mental picture in the physical sphere, is proved by the annoyance which some men show when in spite of great concentration they cannot remember, and which persists even when they have abandoned the attempt to recollect, …

(Aristotle, 1975, pp.303–311)

In the last part of this passage, especially with the reference to frequency, Aristotle is distinguishing between natural and customary associations. Earlier in the passage, when he mentions *loci* he is referring to the ancient "artificial" art of memory by associating, for example, people with given places. Ancient and medieval texts are full of a wonderful range of examples of such use of spatial places as an aid to memory. Aristotle does not use a Greek term for association, but it is implied in phrases such as "pass rapidly from one step to the next" or when he says slightly earlier than the quoted passage "Arts of recollection occur when one impulse naturally succeeds another" (p.301). Finally a few lines later on the same page he describes what are sometimes called his three laws of association.

This is why we follow the trail in order, starting in thought from the present, or some other concept, and from something similar or contrary to, or closely connected with, what we seek. (Aristotle, 1975, p.301)

Here similarity is just like Hume's resemblance, and "closely connected" with contiguity.

The maxims and heuristics of the ancient art of artificial memory were aimed at the facilitation of memory, but the associations used, often with an emphasis on vivid and striking images, are not far removed from those Freud encountered in the free associations of his patients. (For the history of the art of memory, see Yates, 1966.)

To make another point, I want to say something more explicit about what I mean by *free* associations, since the general theory of associations

covers a large part of executing practical activities. In such activities the associations are not free, but conditioned in a fixed sequence to accomplish the task at hand. As the standard phrase goes, they have become automatic. Free associations are of a different sort, consciously used in memory searches, for example, when automatic retrieval is not working. Free associations are more characteristic during moments of meditation or reverie, but also as unexpected intrusions of images unrelated to the task at hand, prompted by any of a great variety of possibilities.

The message I am trumpeting is that of learning to recognize the guidance and the help we can get from such associations, and perhaps even more, from those that do not rise to consciousness, but that are expressed in action by our actual choices. We often describe such choices as instinctual, as "the one I liked but I can't say why," or as "the one that seemed familiar but I can't explain it."

Free associations are a mixed bag, some come with positive affect and some not. A good example of "not" is to be found in the early pages of Joyce's *Ulysses* (1934, pp.7–11) as Stephen Daedalus ruminates about the death of his mother following Buck Mulligan's remark that he killed her by his stubborn refusal to kneel and pray at her bedside as she lay dying. Such inward-turning ruminations can interfere with the quality of associations and thus of choices. Experimental confirmation of this claim is to be found in Wilson and Schooler (1991) and related studies referred to there.

The variety of empirical studies that I would classify as relevant to the understanding of free associations is very large. But there are two broad, not quite orthogonal, classifications of the most importance. One is the distinction between those having positive or negative affect, and the other is between being inward or outward directed. The connections between ruminative, negative-affect associations and psychological depression have been much studied. The detailed complex conclusions cannot be summarized here, but a good overview is to be found in Nolen-Hoeksema (1991).

Even though I am persuaded that the theory of rationality, or of freedom for that matter, in the fullest sense should include the psychological concepts and problems mentioned in the preceding paragraph, it is not feasible to go further here. I do think there has been far too much separation between the conceptual approaches to choice behavior of economists, on the one hand, and social or personality psychologists, on the other. Only in the empirical studies of consumer behavior have we as yet seen a real reduction of this separation.

A formal remark on utility. Even though I am, as already expressed, skeptical of the grander schemes of how expected utility is maximized in the choices of ideally rational persons, there is a natural connection between the probabilistic phenomena of free associations in choice and random utility models. Much of the current literature on choice in the social sciences, especially economics, uses random utility functions (e.g., McFadden & Train, 2000). A strict derivation of such utility models from basic assumptions about momentary mental associations is straightforward. We sketch the mathematical argument here. We define the momentary random utility of a choice response r at time t by the sum of the strengths of momentary associations of the brain image of r to brain images of scenes, pictures, persons, and so forth at t. The decision rule is now that a choice response r is made at time t if its momentary random utility at t is the maximum among choice responses available. In other words, among the possible choice responses, the probability of choice response r occurring is just the probability that at time t the momentary associative strength of the brain image of r is the largest. The essential connection of utility to free associations is made by using fluctuating random utility functions, which literally change from moment to moment, as advertisers realize and depend on to promote their products.

2.4. Psychological nature of the verification of informal mathematical proofs (Suppes, 2005)

Even less than what was said in the first example on truth computations can be easily said about the intuitive steps, without explicit formal verification, in informal proofs. But I would defend the proposition that in such proofs we continually use patterns of associations that are more complicated and subtle than those needed in my truth examples. Yet I suggest, it is a feasible psychological project to survey the main features of such patterns in the informal proofs that occur in a given area of mathematics. Memory of many such patterns is undoubtedly a mark of being an expert in a given domain. Perhaps even more important is having a feeling of how to judge correctly the similarity of a prior pattern, widely held to be valid, to a new one being evaluated. Such experienced judgments of similarity are not at all special to proofs, but occur in every area of experience from case studies of the law to athletic skills of every variety. The content is special to the domain, but the general empirical character is not. (For an introduction to the formal aspects of the large psychological literature on similarity,

often an intransitive relation due to thresholds, see Chapter 14 on proximity spaces and Chapter 16 on representations and thresholds of Suppes, Krantz, Luce, and Tversky (1989), which also contains extensive references.)

What I have said is too general, but can quickly be extended to more specific considerations by examining some examples of informal proofs. For reasons of space, I restrict myself to two.

Example 1. Proof of an Archimedean axiom, taken from Royden (1963). The axiom *C* referred to in the proof is the standard completeness axiom: every nonempty set *S* of real numbers which has an upper bound has a least upper bound. Here is the theorem and informal proof, as given by Royden:

> **Axiom of Archimedes**: *Given any real number x, there is an integer n such that x < n.*
>
> *Proof*: Let *S* be the set of integers *k* such that $k \leq x$. Since *S* has the upper bound *x*, it has a least upper bound *y* by axiom *C*. Since *y* is the least upper bound for *S*, $y - \frac{1}{2}$ cannot be an upper bound for *S*, and so there is a $k \in S$ such that $k > y - \frac{1}{2}$. But $k + 1 > y + \frac{1}{2} > y$, and so $(k + 1) \notin S$. Since $k + 1$ is an integer not in *S*, we must have $k + 1$ greater than *x* by the definition of *S*.
>
> (Royden, 1963, p.25)

As expected, there is no filling out of obvious simple arguments. For example, "$y - \frac{1}{2}$ cannot be an upper bound for *S*." The formal expansion is obvious but tedious. What is important here, and critical for informal proofs, is the power of ordinary language along with a minimum of notation to describe the argument that could easily be written as an algorithm. In saying this I am not claiming that we know how to write general algorithms for any such gaps. In general form they may not exist, because of well-known undecidability results, or high lower bounds on such decision procedures as Tarski's for the first-order theory of real closed fields. The last sentence of the proof exhibits a similar use of ordinary language to summarize informally the argument.

Example 2. This one concerns equivalents of the axiom of choice. I take the example from my own book on axiomatic set theory (Suppes, 1960/1972). A useful maximal principle, due independently to Teichmüller (1939) and Tukey (1940), is characterized by defining when a set is of *finite character*, which is true of a set *A* if and only if

(i) *A* is a nonempty set of sets,
(ii) every finite subset of a member of *A* is also a member of *A*.

The intuitive idea behind this formulation is that a property is of finite character if a set has the property when and only when all of its finite subsets have the property.

> **Teichmüller-Tukey Lemma**: T. *Any set of finite character has a maximal element*.

The theorem of interest is:

> *Theorem*: *The Teichmüller-Tukey Lemma* T *is equivalent to Zorn's Lemma Z*.

Recall that Zorn's Lemma states that if $A \neq 0$ and if the sum of each nonempty chain which is a subset of *A* is in *A*, then *A* has a maximal element, where *A* is a chain if and only if *A* is a set of sets and for any two sets *B* and *C* in *A* either $B \subseteq C$ or $C \subseteq B$.

> *Proof*: We prove only the first half, namely, that Zorn's Lemma (*Z*) implies the Teichmüller-Tukey Lemma.
>
> Let *A* be a set of finite character, and let *C* be any chain which is a subset of *A*. To apply *Z* we need to prove that $\cup\, C \in A$. Let *F* be a finite subset of $\cup\, C$. Then *F* is a subset of the union of a finite collection *D* of members of *C*, for each element of *F* must belong to some member of *C* and there are only a finite number of elements in *F*. Now since *D* is finite and is a subset of the chain *C*, it has a largest member, say *E*; and *F* must be a subset of *E*, for otherwise *C* would not be a chain. $E \in A$, whence since *A* is a set of finite character, $F \in A$; but then also $\cup\, C \in A$. The hypothesis of *Z* is thus satisfied by *A* and by virtue of *Z*, *A* has a maximal element.
>
> (Suppes, 1960/1972, p.249)

A first rough comparison to the length of this informal proof to formal ones that assumed the same background of prior theorems may be made using empirical data in Suppes and Sheehan (1981, p.79) on the length of nine formal proofs made by students in a course I taught for many years on set theory, for which a computer-based proof checker was developed and then regularly used. (Details are in the article just cited.) The mean length, is terms of number of lines with explicit inference rules,

was 41.8, with the min = 25 and the max = 61. As expected, the informal proof given above is much shorter.

The central characteristic of this informal proof, like the previous one, is the use of ordinary English sentences with some embedded mathematical symbols to summarize intuitively individual arguments, each of which correspond approximately to a number of steps in a formal proof.

A psychological point about this linguistic feature of many informal written proofs is the implication for understanding such sentences. Undoubtedly the problem of being satisfied with a personal verification of an informal proof is quite dependent on the intuitive mathematical clarity of the written form of the informal proof. So, often it is not the overall structure of the proof, but the difficulty of comprehending individual sentences. To comprehend such sentences the reader often needs to be able to build a mathematical model satisfying the sentence, and often some other sentences as well as visual graphs and the like, at least in sufficient detail to feel the model is enough. Mathematicians are good at this. It is an essential part of what they have learned. Intellectually, this differs from algorithmic checking in a way that is parallel to the difference between model theory and proof theory.

From a broader perspective, the contrast between formal and informal proofs is striking. A formal proof manipulates symbols, and to check such a proof a computer program needs to have no understanding at all of the symbols to make an evaluation of correctness. A corresponding informal proof ordinarily does not mention mathematical symbols or language of any sort, but only mathematical objects, numbers, not numerals, operations on sets, not the notation for the operations, etc. The linguistic demand, in this case, is semantically driven, not syntactically. Understanding of the informal language used to talk about nonlinguistic mathematical objects is necessary.

Finally, I want to emphasize that the model sketches, as I am calling them, used to check informal proofs, are special to mathematicians only in part, not in their general psychological features, which are surely shared by architects, builders, and designers of all kinds who rely on a variety of images, externalized on paper or on a computer screen, but also images of the imagination, to facilitate thinking about whatever problem is current.

On the basis of the sound methodological principle that properties of the mind are really properties of the brain, it is useful to see what can be said about informal proofs from the standpoint of brain activity, even though it is obvious we have at present many mental concepts we cannot characterize in terms of what we know about the brain.

The first observation is one that brings us back to formal proofs. It is widely recognized by almost everyone working on the subject that all computations are physical computations. In other words, any actual computations require a physical embodiment. This does not mean that digital computers are in any sense the universal model of how computations are made. Natural computations in the biological world have an endless variety of physical embodiments. The computational nature of DNA as the genetic code is probably the greatest single scientific discovery of the second half of the twentieth century. But the problems of understanding the physical computations of the seemingly simple motions of the thousands of insect species are overwhelming in their complexity and diversity. How, for example, does an ordinary house fly compute its escape route of flight from a detected predator?

Formal proofs are certainly recognized as having relatively easy implementation as physical computations on a digital computer. Whatever abstract talk there is about the meaning and implications of a formal proof, the verification of the proof is a recursive physical process, painfully explicit in its details, as emphasized earlier from a different perspective.

Something similar has to be true of informal proofs, with brain computations replacing digital ones. A much too simple model of such brain computations was given earlier in the analysis of how the truth is computed of ordinary empirical statements about highly familiar matters, such as the most obvious sort of geographic or demographic facts. At the psychological level, the method of computation proposed is that of association. Informal proofs are a triumphant application. At least until recently, many psychologists unfamiliar with the detailed analysis of mechanisms of computations were inclined to be skeptical of such a claim. But the general complexity analogy with digital computers is too obvious to tolerate any wholesale rejection of association as the primary basis of the brain's computations. Yet, as I have already stated, it is a long way from Minsky's simple universal Turing machine (1967) with four symbols and seven internal states to the complexity of a digital computer with programs able to

170

defeat the best human chess players, for example, or make a trillion computations to predict the weather. So it is with the brain, from the simple associations of small invertebrates like *Aplysia* to the most intricate mathematical proofs.

I have introduced, in these last paragraphs, many inadequately developed ideas about how the brain works. Full details are available nowhere, but will undoubtedly be a subject of intense research for many years to come. There is not space to try to say in a careful way what I think we do, at the present, know. So I will end with two remarks, one speculative and one empirical.

The first remark concerns meaning. It is a standard complaint of many years that Hilbert's formal systems for the foundations of mathematics turn mathematics into a meaningless game. The arithmetic of numerals, as opposed to numbers, has nothing like the rich content of genuine number theory or geometry, a source of endless intuitions and meaningful relations. The new home of Hilbert-style formalisms is, of course, in computer science, and much more broadly, the programming efforts throughout the world to use formal languages to write computer programs, which implement solutions to a vast array of tasks and problems. There is no additional sense or meaning given to the computer as part of these programs. It is formalism all the way down, but with physical embodiment. Moreover, it is hard to think of its being any other way.

Detailed thinking about the brain moves in the same direction. There is no mysterious Fregean sense lurking somewhere in the cortex, ready to supply meaning as needed. The meaning of a word, a phrase or a sentence, like the meaning of a perceptual image, is to be found in a welter of associations, or, to put it more soberly, in associative networks that are, in humans, if not in *Aplysia*, of great complexity. Of course, to put it this way is too bald and simple, as if reference to complexity were sufficient to explain how a predictive model of the weather computes an estimate of what the weather will be like the day after tomorrow. It helps not at all, in concrete terms, to say the program uses a terabyte of memory and computes at the rate of two terabits a second. Sustained research will be required for the indefinite future to untangle just how the brain is computing any important task, but the associative nature of the computations is, on present evidence, a reasonable conjecture.

Displacing the Aristotelian and Cartesian conceptions of mind, attractive and empirically sound as they were in many respects, will, when fully accomplished be comparable in intellectual importance to the dis-

placement of Ptolemaic astronomy by that of Copernicus, Kepler and New-ton. This philosophical and scientific revolution of the mind was given a big boost by Hume's *Treatise of Human Nature* (1739/1888). Moreover, the claim that the workings of mind reflect above all the workings of collections of synchronized neurons linked in a complex associative network is broadly accepted already in neuroscience, and much of psychology, but by no means to the same extent in philosophy.

The second and final remark is more down to earth and empirical. An early question that arises in thinking about how the brain processes ordinary language, including that used in informal proofs, is, how does the brain process linguistic input, i.e., language for listeners? (The question of linguistic output, i.e., speech production, is even more complicated.) Here I have in mind the relatively simple question of just what can we say about the initial brain processing of words and sentences heard. (I consider here only spoken language, but most of what I have to say applies just as well to the visual process of reading.) Now the analysis of sentences as finite sequences of words, and the analysis of words as sequences of syllables, and, at the more detailed level, sequences of phonemes, is widely accepted as being approximately correct. In addition, for many reasons, it is a sensible hypothesis to expect that the methods of brain computation are likely to preserve approximately the temporal order of words in sentences, syllables in words, and so forth.

To test this idea in brain data is to test a hypothesis of structural isomorphism between spoken sentences and their brain representations. I summarize some unpublished work with my younger colleagues (Suppes, Perreau-Guimaraes, and Wong, In press). Sentences are presented auditorily or visually, one word at a time on a computer screen, at the temporal pace of the auditory recordings. Electric waves in the cortex, time-locked to the presentation of each sentence, are recorded for each subject using standard electroencephalography (EEG) techniques. Various linear models, such as those based on Fourier transforms and filters, or one-layer neural networks, are used to eliminate noise and find an approximately invariant signal (Wong, Perreau-Guimaraes, Uy, and Suppes, 2004). The measure of success at the first level is being able to classify correctly a significant number of test trials not used in estimating the parameters of the model being evaluated. The sentences were of the geographic type mentioned earlier, and the subjects were asked to judge each one as true or false, and so indicate by typing '1' for true and '2' for false. Good, but far from perfect, recognition results were obtained by sets of sentences of size 24, 48 or

100. We also isolated in the temporal sequence of presentation, individual words to which the same models were applied.

Let f be a one-one function mapping each sentence s to its brain representation $f(s)$. Let g be a corresponding function for words. Then our test of structural isomorphism is whether or not we can find empirical support for the structural equation, where $s = w_1\, w_2\, \dots\, w_n$,

$$f(w_1\, w_2\, \dots\, w_n) = g(w_1)\, g(w_2)\, \dots\, g(w_n).$$

It perhaps seems too obvious that this equation should hold, just by asking how else could the brain process sentences. But already at the level of speech such precise identification is not always easy, so it is a serious problem whether or not such results can be substantiated in the brain. We have good support, but not as good as our recognition of sentences, which is not surprising, since this same relation of relative difficulty holds for speech.

Finding support for such a natural isomorphism seems necessary to get started, but it is clearly a long journey of further results to get to such questions as how informal proofs are processed in the brain. In this case, much more than the initial isomorphism of recognition is needed. Semantic computation in the spirit of the associative networks and model sketches mentioned earlier are essential. Still, at a certain level the task is well-defined, and I see, at the present, no alternative conception that is more promising. Whatever empirical route does prove successful, I find it unimaginable that there can be a fully satisfactory theory of the verification of informal proofs that is *a priori* and devoid of psychological, and ultimately neural, concepts and data.

REFERENCES

Aristotle (1975). On memory and recollection. In G.P. Goold, ed., *On the Soul [De Anima], Parva Naturalia, On Breath*, pp.287–315. Cambridge, MA: Harvard University Press, 4th edn. English translation by W.S. Hett. First published in 1936.

Blackburn, P. and J. Bos (2005). *Representation and Inference for Natural Language: A First Course in Computational Semantics*. Stanford, CA: CSLI Publication.

Burt, C.L. (1925). *The Young Delinquent*, 4th edn., London: University of Toronto Press.

Collins, A.M. and E. Loftus (1975). A Spreading Activation Theory of Semantic Processing. *Psychological Review*, 82, pp.407–428.

Davis, J. A. (1964). *Great Aspirations: The Graduate School Plans of America's College Students*. Chicago: Aldine.

de Finetti, B. (1937/1964). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7, pp.1–68. Translation in Kyburg and Smokler (1964).

Freud, S. (1971). New Introductory Lectures on Psycho-analysis, and Other Works. In J. Strachey, ed., *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Vol.XXII*, p.11. London: Hogarth Press. First published in 1964.

Gaudet, H. (1955). A Model for Assessing Changes in Voting Intention. In P.F. Lazarsfeld and M. Rosenberg, eds., *The Language of Social Research*. New York: Free Press of Glencoe.

Ghilesin, B. (1952). *The Creative Process*. New York: Mentor.

Goode, W.J. (1956). *After Divorce*. New York: Free Press of Glencoe.

Hadamard, J. (1945). *The Psychology of Invention in the Mathematical Field*. Princeton: Princeton University Press.

Hume, D. (1739/1888). *A Treatise of Human Nature*. Oxford: Oxford University Press. Selby-Bigge edition.

James, W. (1890/1931). *Principles of Psychology*, Vol.I. New York: Henry Holt.

Joyce, J. (1934). *Ulysses*. New York: Random House.

Kant, I. (1781/1997). *Critique of Pure Reason*. Translated by P. Guyer and A.W. Wood. New York: Cambridge University Press.

Kornhauser, A. and P.F. Lazarsfeld (1955). The Analysis of Consumer Actions. In P.F. Lazarsfeld and M. Rosenberg, eds., *The Language of Social Research*. Glencoe, IL: Free Press.

Lazarsfeld, P.F., ed., (1931). *Jugend und Beruf*. Jena, Germany: Fischer.

Maier, N.R.F. (1931). Reasoning in Humans: II. The Solution of a Problem and its Appearance in Consciousness. *Journal of Comparative Psychology*, 12, pp.181–194.

McFadden, D. and K. Train (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, 15, pp.447–470.

Minsky, M.L., (1967). *Computation: Finite and Infinite Machines*. Englewood Cliffs, N.J.: Prentice-Hall.

Murphy, G.L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Nisbett, R.E. and T.D. Wilson (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84, pp.231–259.

Nolen-Hoeksema, S. (1991). Responses to Depression and their Effects on the Duration of Depressive Episodes. *Journal of Abnormal Psychology*, 100, pp.569–582.

Proust, M. (1927/1999). *Time Regained: In Search of Lost Time*. Translated by A. Mayor and T. Kilmartin. New York: Modern Library.

Rossi, P.H. (1955). *Why Families Move: A Study in the Social Psychology of Urban Residential Mobility*. New York: Free Press of Glencoe.

Royden, H.L. (1963). *Real Analysis*. New York: The Macmillan Company.

Savage, L.J. (1954). *Foundations of Statistics*. New York: Wiley.

Simon, H. (1955), A Behavioral Theory of Rational Choice, *Quarterly Journal of Economics*, 69, pp.99–118.

Suppes, P. (1956). The Role of Subjective Pprobability and Utility in Decision-making. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955*, 5, p.72.

Suppes, P. (1960/1972). *Axiomatic Set Theory*. New York: Van Nostrand, slightly revised edition published by Dover, New York, 1972.

Suppes, P. (1969). Stimulus-response Theory of Finite Automata. *Journal of Mathematical Psychology*, 6, pp.327–355.

Suppes, P. (1977). Learning Theory for Probabilistic Automata and Register Machines. In H. Spada and W.F. Kempf, eds., *Structural Models of Thinking and Learning*, pp.57–79. Bern: Hans Huber Publisher.

Suppes, P. (2002). *Representation and Invariance of Scientific Structures*. Stanford, CA: CSLI Publications.

Suppes, P. (2003) Rationality, Habits and Freedom. In N. Dimitri, M. Basili and I. Gilboa, eds., *Cognitive Processes and Economic Behavior*. Proceedings of the Conference held at Certosa di Pontignano, Siena, Italy, July 3–8, 2001, pp.137–167. Routledge Siena Studies in Political Economy. New York: Routledge.

Suppes, P. (2005). Psychological Nature of Verification of Informal Mathematical Proofs. In S. Artemov, H. Barringer, A.S. d'Avila Garcez, L.C. Lamb, and J. Woods, eds., *We Will Show Them: Essays in Honour of Dov Gabbay, Vol.2*, pp.693–712. London: College Publications.

Suppes, P. (2007). Where do Bayesian Priors Come from? *Synthese*, 156, pp.441–471.

Suppes, P. and J.-Y. Béziau (2003). Semantic Computations of Truth, Based on Associations already Learned. *Journal of Applied Logic,* 2,pp. 457–467.

Suppes, P., B. Han, J. Epelboim, and Z.-L. Lu (1999a). Invariance between Subjects of Brain-wave Representations of Language. *Proceedings National Academy of Sciences*, 96, 12953–12958.

Suppes, P., B. Han, J. Epelboim, and Z.-L. Lu (1999b). Invariance of Brain-wave Representations of Simple Visual Images and their Names. *Proceedings National Academy of Sciences*, 96, 14658–14663.

Suppes, P., D.H. Krantz, R.D. Luce, and A. Tversky (1989). *Foundations of Measurement, Vol.II. Geometrical, Threshold and Probabilistic Representations*. New York: Academic Press.

Suppes, P., Z.-L. Lu, and B. Han (1997). Brain-wave Recognition of Words. *Proceedings National Academy of Sciences*, 94, 14965–14969.

Suppes, P., M. Perreau-Guimaraes, and D.K. Wong (In press, Neural Computation). Partial Orders of Similarity Differences Invariant between EEG-recorded Brain and Perceptual Representations of Language.

Suppes, P. and J. Sheehan (1981). CAI Course in Axiomatic Set Theory. In P. Suppes, ed., *University-level Computer-assisted Instruction at Stanford: 1968–1980*, pp.3–80. Stanford, CA: Stanford University, Institute for Mathematical Studies in the Social Sciences.

Teichmüller, O. (1939). Braucht der Algebraiker das Auswahlaxiom? *Deutsche Mathhematik*, Vol.4, pp.567–577.

Themistius (1966). *On Aristotle's "On the Soul"*. Translated by Robert B. Todd. New York: Cornell University Press.

Tukey, J.W. (1940). *Convergence and Uniformity in Topology*. Annals of Math. Studies, No.2, Princeton.

Wilson, T.D. (1985). Strangers to Ourselves: The Origins and Accuracy of Beliefs about One's Own Mental States. In J.H. Harvey and G. Weary, eds., *Attribution: Basic Issues and Applications*, pp.9–36. Orlando, FLA: Academic Press.

Wilson, T.D. and J.W. Schooler (1991). Thinking too much: Introspection Can Reduce the Quality of Preferences and Decisions. *Journal of Personality and Social Psychology*, 60, pp.181–192.

Wong, D.K., M.P. Guimaraes, E.T. Uy, and P. Suppes (2004). Classification of Individual Trials Based on the Best Independent Component of EEG-recorded Sentences. *Neurocomputing,* 61, pp.479–484.

Yates, F.A. (1966). *The Art of Memory*. Chicago: University of Chicago Press.