UNIVERSITY OF CALGARY

Psychological Empiricism

by

James Cherry

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF ARTS

DEPARTMENT OF PHILOSOPHY

CALGARY, ALBERTA

June, 1999

0-612-47935-8

It's lovely to live on a raft. We had the sky, up there, all speckled with stars, and we used to lay on our backs and look up at them, and discuss whether they was made, or only just happened - Jim he allowed they was made, but I allowed they just happened; I judged it would have took too long to make so many. Jim said the moon could have laid them; well, that looked kind of reasonable, so I didn't say anything against it, because I've seen a frog lay most as many, so of course it could be done.

Huck Finn

Abstract

This thesis is an attempt at a novel defence of empiricism and anti-realsim. I develop a psychological position which is mechanistic in nature yet still remains consonant with introspection. The position also affords the possibility that the only overlap between the external world and the understanding lies in the association of afferent stimuli. The analog of the association of afferent stimuli in the realm of scientific knowledge is the correlation of observables, i.e. regularities or Humean constant conjunction. The analog of that content of the understanding which is not reducible to the association of afferent stimuli is theoretical confabulation, i.e. the unobservable and metaphysical. The thesis concludes with an explication of the suggested anti-realism and a reply to some common objections to similar positions.

Table of Contents

Behaviouristic ideals, so popular among North American psychologists in the first half of this century, have a certain aesthetic appeal. It has been said that they take the 'person' out of the person. To my mind, the determinism, empiricism, and eschewing of the agent-cause should not be thrown out with the bathwater of behaviourism's inherently circular and vacuous explanations, its exclusive reliance on data driven processes, and its rejection of the importance of mental imagery and goal states. This work represents the start of an effort to develop a psychology which retains the alluring aspects of behaviourism, while at the same time stressing the role of conscious imagery and higher level thinking. The project is to show how a 'scientific' perspective on the mind can be conformable to introspective experience.

Now the empiricism mentioned as a feature of behaviourism is present here in two places. First is the positivistic influence which the writings of Mach and Bridgeman had on Skinner (Skinner 1931), and which Carnap had on Tolman (Leahey 1991). The other sort of empiricism is more Humean. It is the thesis that learning takes place by induction. Though some mechanism of abstraction or generalization has been present in psychology at least since Aristotle, the epistemic consequences of construing all learning in this way should not be underestimated. The discussion of these issues is postponed until the psychological background is in place. Section 6 will deal with the matter in the context of confabulation. Section 7 extends those conclusions to the realm of scientific knowledge. Sections 1 through 5 establish the backgound.

Section 1: Basic Distinctions and Mechanisms

Distinction 1.1: conscious/ imagistic vs. non-conscious system The brain can be divided into two separate systems: the conscious system and the non-conscious system. This division, however crude, is intuitive and straightforward. It derives from the plain observation that some of the workings of the brain are conscious and some are not. By calling them separate systems, I do not mean to imply that the mechanisms which constitute each are subserved by different neurology; nor that either system is indivisible into subsystems; nor even that these systems are wholly distinct.

Fundamentally, there are two things which separate these systems. The first is the fact that the former involves awareness. Awareness, however, comes in degrees. We might call this "strength of imagery" (see section 2). This is what is meant by the denial (of the previous paragraph) that the two systems are completely distinct; in certain respects, we might be better to regard them as lying on a continuum. Metaphorically, we might imagine awareness as a kind of flashlight. The beam can be focused on one area, broadened to cover more ground (at a loss of intensity at each point), or dispersed to focus at several disjoint areas at once. The total intensity is relatively constant and thus limited.

Secondly, these two systems can be (again metaphorically) distinguished according to the 'language' which each employs in its processing. The language of the former is imagistic. That is, propositional/ verbal, visual, auditory, kinesthetic, emotional, and the like: *i.e.*

phenomenological. This fact is evident upon introspection. The language of the latter, by contrast, we might call "neuronal" (for lack of a more descriptive and meaningful term). It is argued that this system is governed by something like ideal laws of conditioning (accounting, somehow, for genetic pre-dispositions). Such processing does not require imagery/ awareness.

The division is for explanatory purposes only. Speaking of interaction between two separate systems affords more intuitive explication.

Distinction 1.2: semantic vs. implicit knowledge

The non-conscious system is regarded as the more fundamental of the two. It probably develops prior to the other, both ontogenetically and phylogenetically, and it controls a wide variety of behaviour. This system governs over all implicit and procedural knowledge (though the conscious system may aid in acquiring certain of these skills and behaviour patterns - see mechanism 4.2): everything from near imperceptible movement to some complex, but non-conscious, strategizing and deliberating. For example, it has been established that operant conditioning can occur without awareness, that the rules of an artificial grammar can be acquired and applied without awareness (Reuber 1989), that the winning strategy for a complex gambling task is discovered and employed before the subject even realizes that she is employing any strategy at all (Bechara *et al.* 1997). Similarly, the non-conscious system is capable of perception, as demonstrated by the cocktail party phenomenon as well as blindsight.

Semantic knowledge, on the other hand, relies on the conscious system for its acquisition and application. This type will be detailed in the next section.

Mechanism 1.1: non-conscious learning - conditioning

As stated, the non-conscious system is probably governed by something close to the laws of conditioning. In effect, conditioning is a kind of trial and error process. The organism acquires associations between circumstances and action (really between afferent stimuli) by 'randomly' emitting operants, then repeating that which is reinforced and discontinuing those which are not. When a winning strategy is found the organism persists in employing it, when the organism loses, new behaviours are sought. Now the organism or non-conscious system is not always relegated to a random search for appropriate operants. First, the circumstances may bear a similarity to something previously encountered, and thus similar behaviours will be emitted. Second, in some cases the genetic predispositions of organisms will dictate just what kinds of behaviours are emitted, and what kinds of associations can be acquired. However, for artificial or completely novel situations, conditioning does approach a real trial and error procedure. Either way, the notion of "trial and error" is not meant to stress the random nature of operant production, but rather, the hypothesis testing strategy of conditioning.

It is easy for this system to learn such things as the fact that a switch will turn on the light, or that fire will burn and cause pain. There are, however, tight constraints on the circumstances which are necessary for this type of learning to occur. For example, the time delay between the operant and the reward, or the conditioned and unconditioned stimuli, must not exceed a few seconds. Also, there are constraints on the covariation between the relevant events in that it must be very high, at least at first, for conditioning to occur (Ferster and Skinner 1957). As such, it would be impossible for anyone to learn a large part of the knowledge we all have, and apply regularly, by conditioning alone. For example, it would be impossible to learn that coffee has a stimulating effect, or that exercise makes one stronger, or even that intercourse causes pregnancy. This kind of learning relies on the conscious system, and takes place according to mechanisms other than conditioning. Such semantic knowledge relies on the relations between images (mostly verbal in this case) which constitute the schemata.

Section 2: Schemata

The difference in language of processing between the two systems might well be accounted for in terms of organization. The conscious system stores information into schemata. These, effectively, are the knowledge structures which allow for groupings and relations of diverse, sometimes simple images.

For example, words may be grouped semantically, phonetically, by the first letter or last syllable (rhymes). Conrad (1964) found that on immediate recall of visually presented letters, a misremembered "B" is more likely to be recalled as a "V", which is phonetically similar, than as an "R", which is visually similar. Similarly, Bousfield (1953) observed that a memorized list of items, each of which belong to one of four categories, are recalled by category despite the random order of presentation. This organization facilitates the recall of, for example, words which rhyme with "orange", towns which start with the letter "F", or things which are small, furry, and say "meow".

Schemata also enable images to be related, in various ways, to images in other modalities. For example, the word "house" can be a verbal image, and related to the articulatory-kinesthetic image of the word, a visual image of "my house" or an abstracted house, *etc.* Also, since the relations between images can be causal, (and temporal, spacial, superordination, similarity, *etc.* - in fact, any imaginable relation can tie two images, indeed by definition), schemata can account for our causal theories (which play a major role in memory and behaviour determination), as well as our timeline and spacial organizations of reality.

In the previous section, it was said that schemata were to explain semantic knowledge. We can now see how the relations between distinct simple images can account for causal theories such as: the ingestion of coffee causes stimulation so that sleep becomes more difficult. We can imagine causal theories as stored in discrete images related in various ways. For example, this theory might be constituted by the distinct images: coffee, caffeine, ingestion, stimulation, and sleep. Plus the relations between these: caffeine 'is contained in' coffee, caffeine 'causes' stimulation, stimulation 'inhibits the ability to' sleep, *etc*. The same process of organizing semantic knowledge into distinct images and relations should be able to account for the entire content of verbal imagery.

It is evident that the distinct images of the preceding example are far from simple. Indeed, it may be argued that a distinction between simple and complex imagery is untenable. Though that is probably true, the concept of simple images affords an analogy useful for explanatory purposes. The distinction will be discussed more thoroughly in section 5.

Mechanism 2.1: schematic learning 1 - cultural transmission

There appear to be two ways in which schemata can be learned, or images and relations acquired. The first is that they are taught. This is the way that cultural theories are transmitted through generations. Everything from religious theory to the optimal time to plant corn, to the laws of morality, to mathematics. A mother teaches her child never to go near the stove, or to use sunscreen in the summer. (Contrast this with the mechanism of conditioning whereby the child learns not to touch the stove by getting burned on some occasion.) This method of learning involves the forming of new associations or relations between previously existing images, and the creation of new images thereby.

Mechanism 2.2: schematic learning 2 - self-acquisition

Theory and semantic knowledge can also be acquired without being taught, but by discovery/ invention. Indeed this type of learning is necessary to explain both the invention of what becomes cultural knowledge, as well as knowledge which is not shared (*i.e.* private knowledge or episodic memories). This mechanism takes place by hypothesizing and testing hypotheses. The hypotheses are generated by searching the existing schematic network for a plausible cause (mechanism 5.1). The testing process takes place through something like the interpretation of experience or observation by the theory (see mechanism 2.3).

It is interesting to note in this context that people are actually quite poor at generating new theories independently, and at recognizing or even searching for evidence which refutes existing theories (Nisbett and Ross 1980).

One particularly amusing and instructive example comes from B.F. Skinner's repertoire of experiments. He observed that food deprived pigeons would be conditioned despite the fact that their behaviour was not in any way correlated with the probability of reward. He labelled this phenomenon "superstition". The experimental setup involved the usual 'Skinner box', except that the schedule of reinforcement was fixed: the food magazine would appear every fifteen seconds regardless of the bird's behaviour. In effect, the pigeons would happen to be engaged in some random action when the reward was presented. The behaviour was reinforced and hence continued, becoming stronger with each reinforcement. The cycle repeated until he had his birds turning counter-clockwise, swinging their heads, and pecking at the floor.

> The bird behaves as if there were a causal relation between its behaviour and the presentation of food, although such a relation is lacking. There are many analogies in human behaviour... A few accidental connections between a ritual and favourable consequences suffice to set up and maintain the behaviour in spite of many unreinforced instances. (1948 p.171)

Though this is an example of conditioning and not schematic learning, the parallels run deep. The idea of hypothesis testing or trial and error will become more important as it becomes more explicit throughout this paper. For now, the point is simply that the superstition of the pigeons is not far removed from actual human behaviour. Later we will see the mechanisms which justify that claim. They involve the top-down, conceptual and expectancy driven processes which distort reality and can result in confirmation biases and belief perseverance similar to this superstitious behaviour.

Mechanism 2.3: translation/ attribution and introspection

The process by which non-conscious neural impulses become conscious or imagistic is a process of translation, integration and gap-filling. Translation in the sense that the input into the conscious system undergoes a reorganization by, and into, one of the forms or schemata discussed above: the one with the best 'fit'. In effect, the impulses are categorized and labelled: schematically organized. In this way, impulses may be translated into appropriately grammatical and logical verbal imagery, or into familiar visual images (or kinesthetic, olfactory, etc.).

A special case of this is the translation of perceptual impulses which, in imagistic form, have objects and events as content. That is, what we consciously perceive are things like cars, houses and people, rather than retinal patterns or arrays of coloured pixels. The content of conscious awareness (imagery) is always schematically organized; this organization is a necessary condition for awareness.

The process is also integrative in the sense that impulses from various sources (perceptual and stored knowledge, for example, or perceptual impulses in different modalities) are brought together and put into context. This integration may result in the well known unity of conscious experience.

Gap-filling occurs when the initial impulse does not have an exact fit, or when it is incomplete (either due to lack of relevant knowledge or lesioning). This aspect can explain certain cases of provoked confabulation in brain damaged individuals, as well as the confabulatory/ inferential nature of the causal explanations offered by healthy people (Joseph 1986, 1996).

In effect then, the non-conscious impulses are converted and organized into schematic or imagistic form. An event might receive a temporal and spatial designation, as well as a causal interpretation ("Why did it happen?" or "Why did I do that?"), just as an object is recognized/ interpreted/ translated as "One of those". Importantly, the application of this mechanism can range from purely data-driven to fully expectation-driven. For example, one stares at one of those funny pictures at the mall. Eventually, the randomness disappears and is replaced by a three-dimensional image. This is data-driven translation. By contrast, one attributes the fact that it is raining to their sacrifice of the goat and their expertise at rain-dancing. This is conceptually-driven attribution.

Let us examine four special cases of this mechanism in more detail. Translation/ attribution is the process by which such things as the causes of behaviour (*i.e.* reasons for our actions), attitudes toward people, activities, types of events *etc.*, and emotional and other inner states become 'known' to the agent. Translation also plays an important role in the recall of memories. What is common to these four cases (causes of behaviour, attitudes, emotions and memories) is that the process by which such things are discovered is generally called introspection. The 'introspector'/ 'agent' is somehow supposed to reach back into her mind and bring forth that which is sought. The idea is that people have privileged access to their own emotional states and decision making processes. It is as if the reasons for our actions, memories *etc.*, are simply there, waiting to be asked. The key to this mechanism is that these 'facts' are not simply recalled, but attributed based on data from sources other than introspective access. As each of the four examples is explicated, it will become clear just how translation is purported to replace introspection.

We begin with the translation of the causes of action. In a classic study of misattribution, Wilson and Nisbett (1977) demonstrated that people are often not privy to the causes of their own behaviour. This particular experiment was conducted in a shopping mall under the guise of a consumer survey. Four identical nylon stockings were displayed under a sign that read, "Institute for Social Research - Consumer Evaluation Survey - Which is the Best Quality?". Subjects volunteered for the 'survey'. Each subject (50 out of 52 of which were female) selected the stocking they thought to be of the highest quality, then were asked by the experimenters for the reason behind their choice. A total of 80 reasons were elicited. (Common responses were "This pair had a better knit/ weave/ sheerness/ elasticity/ workmanship...".) The results demonstrated a pronounced position effect; that is, the further to the right the stocking was, the higher probability of its being chosen as the highest quality. As such, the normative standards dictate that the position of the stocking (and possibly the order examined) had a strong causal influence on the choice. However, not one of the 80 reasons offered was the position; and significantly, most subjects expressed offence and denial upon mention of the possibility that position might have effected their decision. Furthermore, all of the 80 reasons given can be ruled out as causally efficacious since it is a premise of the experiment that all the stockings were identical. (The experiment has been repeated with nightgowns. 378 subjects confirmed these findings (Nisbett and Wilson 1977).)

In this case at least, it appears that the process by which the subjects arrived at a causal explanation for their action was not by introspecting their decision making procedure. If it were, we should expect at least some to offer the position of the pantyhose as cause. Rather, they answer the question "Why did I do that?" by asking "What is a common/ socially acceptable reason for choosing pantyhose based on quality". That is, they attributed a cause to their own behaviour by examining both the action and the circumstances, and by searching their schematic network for an acceptable cause, *i.e.*, a justification. This was also the conclusion drawn by the experimenters. (This question asking and answering procedure is also vital to memory recall and problem solving. It will be discussed again in section 5 under the name "resonance".)

This explanation is further supported by myriad experiments which demonstrate that observer subjects (those which observe or read about an actor's behaviour) tend, in certain circumstances, to attribute the same causes/ motivations for the action, as well as the same internal states (beliefs, attitudes *etc.*) to the actor as the actor himself (Jones and Nisbett 1971, Bem 1972). Since the attributions must necessarily be inferred from overt and observable behaviour (as well as existing knowledge) by the observer, and since the observer tends to produce the same explanation as the actor, we conclude that the actor too infers the cause of her own behaviour, as well as attitudes, from the same data and by the same process (*i.e.* translation).

For example, these results were found in Bem's (1972) simulation of a classic dissonance experiment conducted by Festinger and Carlsmith (1959) (our second case). In the original experiment, subjects performed a variety of somewhat boring and monotonous tasks. Afterward, they were 'hired' by the experimenter for either $1 or $20 to tell another 'subject' in the waiting room (really a confederate) that the tasks were enjoyable and interesting. Finally, the subjects were asked how much they enjoyed the tasks.

The results of the initial study indicate what has become known as the reverse-incentive effect. Subjects in the $1 condition reported that they enjoyed the tasks significantly more than subjects in the $20 condition (whose reports were close to the reports of controls who were not 'hired'). The dissonance theorists hypothesized that the change in attitude obtained in the $1 condition is a result of the motivation to reduce the dissonance between the verbal report and the attitude (Festinger and Carlsmith 1959). Without importing such a drive, the reverse incentive effect can be seen as the result of the inference of the subjects' attitudes based on their perception of their own behaviour, their knowledge of the situation, and their schematic causal theories. Subjects in the $20 condition could justify their action (lying to the next in line) while maintaining their belief that the tasks were boring because they could attribute the lie to the payment. Whereas subjects in the $1 condition (the "insufficient justification" condition) had no socially acceptable justification for the lie. They did, however, have another way to explain/ justify/ attribute their action: by claiming, even convincing themselves, that they did not lie at all, that the tasks were not boring. Said another way, the speech did not provide a rational basis for inferring the subjects' attitude toward the task in the $20 condition because it was seen as motivated by the money. Whereas, in the insufficient justification condition, there is no perceived alternative motivation (lying because a psychologist asked you to seems not to be socially acceptable), so the speech is taken to reflect the subject's actual attitude. Again, if we adopt the idea that the attitude toward the tasks was introspected, then we might have a difficult time explaining the difference in reported attitudes across conditions.

In Bem's simulation, subjects listened to a tape recording describing a hypothetical subject, Bob Downing, and the tasks he supposedly performed. Control subjects were then asked to evaluate Bob's attitude toward the tasks. Experimental subjects then were told either that Bob was paid $1 or $20 to tell the next subject in line that the tasks were enjoyable. All experimental subjects heard the same tape recording of the supposed conversation between Bob and the next subject (no difference in persuasiveness was found across groups in the original experiment). Again, the subjects in the $1 condition rated Bob's attitude toward the tasks to be more positive than those in the $20 condition (reverse-incentive effect). And again, the control subjects' rating of Bob's did not differ significantly from subjects in the $20 condition.

In effect, to infer attitudes and motives we answer questions like "What must my (Bob's) attitude be if I am (he is) willing to behave in this fashion in this situation?". The answer is an attribution based on knowledge of the behaviour, the situation, and learned causal theories (schematic knowledge). Since all this information is accessible to any informed observer, it seems that people have no immediate advantage in determining their internal states and motivations (excepting knowledge of prior history). Consider the words of Daryl Bem as he outlines the core of his "Self-Perception Theory".

> Individuals come to "know" their inner attitudes, emotions, and other internal states partially by inferring them from observations of their own overt behaviour and/ or the circumstances in which this behaviour occurs. Thus, to the extent that internal cues are weak, ambiguous, or uninterpretable, the individual is functionally in the same position as an outside observer, an observer who must necessarily rely upon those same external cues to infer the individual's inner states. (1972 p.2)

Our third case where translation replaces introspection is in the realm of emotion. Whereas we usually regard emotional states as accessible to introspection, Schacter and Singer (1962) have demonstrated the importance of knowledge of circumstances in their determination. That is, translation and attribution play a crucial role in the manifestation of an emotion. This includes both the phenomenological character and the behavioral consequences of the emotional state.

In summary, the experiment involved the injection of epinephrine and the inducement of emotional states by placing the subject in a room with a confederate who acted dramatically euphoric or angry. The experimenters manipulated the subjects' expectations of physiological symptoms by either informing them of the side effects of the injection, or explicitly denying that any side effects would occur. The results were clear. Subjects who expected no physiological symptoms were significantly more susceptible to emotional behaviour and self-reports than subjects who had a perfectly appropriate explanation for their aroused state. That is, whereas the informed subjects were immune to the shenanigans of the confederate, the uninformed attributed their arousal to an emotional state thus acting euphoric or angry and self-attributing that state on a questionnaire. In effect, the emotional states of the subjects were manipulated simply by altering their perception of the circumstances.

> One labels, interprets and identifies this [sympathetic excitation] in terms of the characteristics of the precipitating situation and one's apperceptive mass. This suggests, then, that as emotional state may be considered a function of a state of physiological arousal and of a cognition appropriate to this state of arousal... It is the cognition which determines whether the state of physiological arousal will be labelled as "anger", "joy", "fear", or whatever. (380)

This labelling of patterns of excitation based on aspects of the situation and schematic knowledge is the essence of translation. Though emotional states are usually regarded as something accessible to introspection, the idea that they are attributed should not be completely foreign. This is the experience of not being sure how you feel, or of considering recent history when answering the question "How do you feel?", or even of asking "How should I feel?". This is one place where we might observe translation at work.

Now this is not to say that knowledge of circumstances is the only determiner of emotional states; a pattern of sympathetic activation is also a necessary component. Schacter and Singer do argue, however, that it is possible that the aroused state may be near identical regardless of the specific emotion, and that the cognitions are necessary for the 'raw' state to become the phenomenologically familiar emotions. By analogy, we might say that we learn to 'recognize' emotional states in the same way a mother learns to identify the meaning of her baby's cries.

The final case we discuss where translation replaces introspection is in the realm of long-term memory recall. The exact role of translation in encoding and recall will be detailed in section 3. The relevant point here is to stress the constructive aspect of recall, and show how translation can account for it. One of Bartlett's major intentions in his classic book Remembering was to debunk the analogy of the memory systems as a container of individual traces which could be selected and examined at will. This analogy stresses introspection as the process of recall.

By contrast, Bartlett emphasized construction and rationalization over reproduction. In effect, what he observed was a kind of confabulatory tendency. It is argued that this is the same tendency observed in the attribution of motivations, attitudes and emotional states. That is, translation is an essential component to recall. The important analogy lies in the expectancy. Wherever we see schematic knowledge entering into a translation, we should imagine it as 'pulling forth' from the data a predetermined image. This 'pull' of the conscious system will be made clear when feedback control is discussed. This is what distinguishes top-down from bottom-up or information processing approaches.

Now there are certain aspects of events and experiences which are stored schematically. For example, the place and time, as well as causes and effects. But these represent only certain broad aspects of the memory. The details are reconstructed based on this framework (this is made explicit in section 3). It is this reconstruction which occurs by the process of translation. Details are largely invented according to our existing knowledge of how these events generally occur. Again we see expectancies and preconceptions at work.

We can also see how this mechanism again replaces our ordinary notions of introspection. As introspection is removed from psychology, so goes the need for an 'introspector': the 'agent', scientifically known as the 'executive controller', who commands and wills actions to be performed for certain reasons, who deliberates and decides. This is an idea which is notoriously difficult to analyze, but also one which has been a (sometimes implicit) component in many theoretical systems.

To conclude the discussion of translation, and to understand the process fully, we might trace its development. Vygotsky (1934)has argued that inner speech (verbal imagery/ linguistic consciousness) and external or social speech are distinct structures. His arguments illustrate structural differences, developmental differences (both ontogenetic and phylogenetic), and functional differences. Between the ages of 3 and 7, children exhibit not only social communicative speech, but also egocentric speech. At this stage, thought takes the form of external, egocentric speech. At its peak, egocentric speech comprises almost 50% of the total speech; it progressively becomes internalized, eventually transforming fully into inner speech at approximately age seven. That is, egocentric speech is the developmental precursor to linguistic consciousness (Vygotsky 1934, Joseph 1982). It is the fact that inner speech is vocalized in children which makes egocentric speech an optimal source of objective data on thought and translation.

Generally, egocentric speech is an accompaniment to the child's activities; it is a 'running commentary'. It consists largely in explanations, as the child explains his own behaviour to himself. The main structural difference between egocentric and social speech is the fact that the former is self-directed (Vygotsky 1934, Joseph 1982, 1996) Furthermore, in Vygotsky's experiments, the coefficient of egocentric speech was found to double when the child becomes frustrated, that is, when faced with a problem (p.30). This fact will become important when we discuss the role of imagery in problem solving.

It was found that in younger children, the explanations of activities occur subsequent to the actions; the child self-attributes cause and motivation after the fact. Later, the two occur simultaneously. Finally, toward the end of the egocentric phase, as in the mature adult, the rationalization precedes the action.

The fact that the explanations initially occur after the action suggests that it is based on the child's overt behaviour. This is directly in line with our construal of the process whereby adults attribute motivation for their action. But if the attribution occurs after the fact, how is it that adults can explain their actions before they act? That is, how do we account for the temporal shifting?

I would like to offer the proposal that the child learns to predict his own behaviour through the conditioning of the perceptual impulses, which occur during the action, with impulses which regularly precede that type of action (which we might call "anticipatory impulses").

We might say that the latter impulses contain information about the forthcoming action, but are not yet fully connected to the language areas of the cortex (which subserve verbal imagery) during the egocentric phase. Neural development would thus explain the temporal shifting of the rationalization with respect to the action. Rhawn Joseph (1982, 1996) has offered such an explanation. He argues that the explanations produced in early egocentric speech occur after the action because the left hemisphere has no access to the functioning of the right. He points out that corpus callosal fibres are extremely immature at this stage, and that the fibres do not become completely myelinated until after age 10. Thus, the language areas of the left hemisphere are forced to confabulate an explanation for right hemisphere activity after the fact.

Conversely, we might argue that the impulses themselves do not acquire any 'meaning' for the conscious system until they are associated with specific types of behaviour (that is, with the perceptual impulses which accompany these). This association may be explained by mechanism 1.1 (conditioning) in conjunction with mechanism 2.3 (translation). In this case, what is conditioned are two stimuli: the anticipatory impulse and the perception of overt behaviour. Once the association is made, the former is thereafter translated into schematic form with the appropriate explanations (causal designations). In this way, we come to 'know' what we are about to do by translating the non-conscious impulse into propositional imagery. This proposal can also account for the familiar deliberation procedure (to be detailed in section 4).

The latter explanation is preferred to Joseph's, because his relies on a notion of direct introspective access to both the causes of our action and the forthcoming action itself. Joseph requires the assumption that the impulses themselves contain the relevant information, and all that is required is access to these by the language areas.

In a way then, the attribution of the causes of behaviour always occur after the fact, despite the observation that reasons are often present to awareness temporally prior to the action. It is only because we have experienced that type of action in constant conjunction with the causal designation which is part of the translation that we can predict our own behaviour. The attribution need not be based on the observation of overt behaviour, it may be based on the anticipatory impulses which have been conditioned with the observation of overt behaviour. In either case, reasons may still be considered 'after the fact' in the sense that the succession of images which we know as the familiar deliberation procedure are products of the translation of non-conscious anticipatory impulses (suggestions for action). We do not reason out our course, we inspect and justify suggestions. Once an appropriate action is determined (by mechanisms 4.1 and 4.3) the causal designation of that action becomes our reason. Thus whether or not the justification precedes (temporally) the action, reasons always follow the suggestion. (This is not to say that reasons are not causally efficacious, just that they do not serve the function which is commonly ascribed to them.)

One final brief point needs mention before passing to the concept of strength. It concerns the specific choice of justification/ attribution for some impulses. Many actions are justified by finding a socially respectable reason. We learn (mechanism 2.1) just which motives are acceptable and which are not. In the absence of a mechanism of introspection, the motives cited for any action need not correspond to the actual causes except in so far as the learned causal theory which is applied happens to reflect reality. That is, if one cites the correct cause of his behaviour, it is only by the chance that the causal theory he applied is both correct and appropriate. The mechanism we use to generate our cited motives is driven by forces other than truth; it is largely driven by social factors. "The differing reasons men give for their actions are not themselves without reasons" (Mills 1940).

Definition 2.1: strength of imagery

Strength of imagery is a rather broad concept. Its generality will allow it to serve a wide range of functions. There are at least two ways in which an image can be strong (or weak, or somewhere in between), and these are purported to represent the experience of degrees of awareness.

The first type of strength consists in what we might call "embeddedness". This refers to the situation of an image within the schemata: some are better established in that they have more relations and 'wider' (more often travelled) pathways. In a sense, an image which is stronger in this way is more familiar to the subject. It is this familiarity, in terms of embeddedness or connectedness, which yields the experience of understanding. This type of strength results from efficient organization, and it is this first type which will be most important to the subject of the next section (memory). Also important to efficient memory functioning, images may become better embedded through a process of rehearsal.

Secondly, an image of a certain object or event is stronger in so far as that object or event is present to perception. We label this type of strength "vividness". It is exemplified by the difference in clarity and detail between an image recalled from memory and one translated from a perceptual impulse. This type of strength is especially important in certain cases of problem solving (see section 5). As we will see, imagery sometimes serves as a kind of mental workbench or scratchpad. Because of limited attentional capacity, vividness is crucial to effective problem solving. The vividness of a specific image can depend on such things as word choice. That is, often two phrases might differ in their respective degree of vividness without differing in informational content.

Vividness and embeddedness are subsumed under the same heading of strength because they are not wholly distinct. For example, theoretical concepts which are better embedded will sometimes yield the same kind of advantage for abstract problem solving as perceptual vividness in the realm of practical problem solving.

Section 3: Memory

The main purpose of this section is not to offer an explicit account of the memory systems, but rather to illustrate the role and importance of the conscious system and imagery to these systems. It is argued that the conscious system enables increased memory capacity.

Now, there are two roads to this same point, though they are not really different. We have noted (distinction 1.1) that the main differences between the conscious and non-conscious systems is the language of processing (which was accounted for in terms of organization in section 2), and the awareness which accompanies processing in the former system. Thus, an argument for the advantages of the conscious system in the realm of memory (compared to the non-conscious system alone) will rely on the demonstration of: first, that the schematic organization of impulses yields more efficient storage and recall, and second, that the strength of the imagery is correlated with probability of recall.

Now, it was said that these two arguments are not really different. That becomes obvious when it is realized that strength is meant in the first sense, *i.e.*, embeddedness. Recall that an image is better embedded in so far as it is related to a greater number of other images, and in so far as those relations are themselves stronger. These relations form the essence of schematic organization. That is, the mechanism of translation, the process by which non-conscious impulses become images, can be a process of establishing relations. For example, as a perceptual impulse of an event is translated, it might receive a causal designation (its relations to antecedent or consequent events), a temporal designation (consisting of its temporal relations to events past and future), as well as analogous spacial or similarity or subordination designations ("This is one of 'those' events"). The point is that the organization results from the encoding process which is translation, and that translation consists of establishing relations; schematic organization is really just a process of embedding. Thus the two proposed lines of argument collapse into one. The task at hand, therefore, is to show that (but not how) the use of the conscious system for memory functions yields more efficient storage and increased memory capacity. This will follow from a demonstration that the schematic organization of material yields a higher probability of recall.

It is important to note in this context that it is the relations between various images which constitutes meaning. That is, the meaning of a certain image (say, the verbal image "barber") is completely comprised by its schematic situation, by the particular relations that the image has to other images. Thus meaning is established through organization, and the specific meaning of an image depends on its situation with respect to the whole. This claim will enable the argument that meaning is important to efficient memory functioning to establish our thesis that the conscious system is necessary for the same. This is because meaning is constituted by organization, and the latter is one of the distinguishing features of the conscious system.

The idea that memories can exist independent of meaning has been defended, although it is presently out of fashion. Ebbinghaus studied meaningless memory through the use of nonsense syllables. However, the ecological validity of his results and methodology have been called into question by many people. The current position on the nonsense syllable is that subjects will always apply meaning to the stimulus (Ashcraft 1994). That is, no image will ever stand on its own, unrelated to any other.

Katona's (1940) experiment on the memorization of a series of digits will serve well as an illustration of how organization aids storage and recall. In the first condition, subjects were asked to memorize the digit string: 581215192226. In the second condition, subjects were asked to memorize the amount stated on a card which read: "The Federal expenditures in the last year amounted to $5 812 151 922.26". In the final condition, the card consisted of the same digits in a different configuration: 5 8 12 15 19 22 26. Most of these subjects recognized the pattern fairly quickly (start with 5, add 3, add 4, add 3, add 4...).

On immediate recall, subjects in the third condition did slightly better than subjects in the other two. The more interesting result came on a surprise one week delayed recall task. The first group, not surprisingly, did quite poorly. In the second group, 80% of the subjects recalled that the Federal expenditures exceeded 5 billion, 810 million, though the later digits in the sequence were not recalled. In the third condition, however, recall was almost perfect for the entire series. Even after four weeks, performance of subjects in the third condition exceeded that of the other two groups after only one week. Some who made errors recalled the pattern correctly but started on the wrong number.

The first condition demands rote memorization, something close to a nonsense syllable. The second condition allowed for the use of organization; by introducing meaning to the digits, they could be stored as part of a larger conceptual framework. The only difference between the first two conditions is that in the second, the digits stood for something: the federal expenditure in dollars. This is a comparatively small degree of organization when compared to the third condition, where the digits could be stored as a complete pattern, yet the difference in recall performance was substantial.

Now meaning, construed as organization, comprises more than the concept generally will (and also considerably less than some analyses). It includes all mnemonic devices, from the method of loci and acronyms to causal attribution. In effect, all organizational strategies which aid recall can be seen as analogous.

It should be obvious that mnemonics are simply organizational strategies. Acronyms, peg-word mnemonics and the method of *loci* are all ways of organizing information, of relating that which is to be remembered with that which is easier to remember or that which is already known. Consider, for example, the method of *loci*. This strategy associates a list a stimuli with spacial positions so that imagining the locations serves as a retrieval cue. Now, of course more than just organization is required for this mnemonic. Rehearsal of the associations and vivid visual imagery are also necessary, but these too contribute to strength of imagery.

Bower (1970) describes a study in which subjects learned a list of 40 unconnected items by imagining each in a particular location on a college campus. The list was read aloud once to each subject, the items separated by thirteen second intervals. The familiarity with the campus enabled an average of 38 items recalled immediately, and 34 one day later (largely in the correct order). The method of loci is evidently an extremely powerful mnemonic, for without this strategy, remembering 38 out of 40 items on a list encountered only once is rather improbable. It is argued that the success of the strategy is dependent largely on the organization. As stated, rehearsal and visual imagery are also necessary, but without the relations established between the items and the familiar locations, thirteen seconds of rehearsal (imaging the item itself, unrelated to anything else) would not have allowed for the high level of performance. Thus it seems as though efficient organization cuts down on the amount of rehearsal necessary for an equivalent probability of recall.

Now such mnemonic devices seem to work by organizing the material almost arbitrarily. That is, it should not matter what images are chosen as the locations, or the 'slots' which the items fill, so long as they are easily remembered. But this organization is still meaningful. It still relies on relating or associating new information into existing frameworks. This same pattern is evident in other types of mnemonic devices which are generally considered meaningful, and indeed, not generally considered mnemonic devices at all. Causal attribution is an example. The analogy is as follows: the method of *loci* involves organizing material into a pre-existing spacial framework which serves as a retrieval cue, causal attribution involves organizing events into pre-existing casual schemata which serves as retrieval cue.

Edwards and Potter (1992) argue that the memories of events are constructed around a framework of causal attribution. That is, we construct memories in order to bolster or rebut certain versions of events, to suggest certain inferences about motivations, dispositions, and personality, as well as to attribute responsibility, praise or blame.

> Ideas about cause and motivation provide the kinds of narrative and inferential links that are the basis of both memorability (providing coherence which aids recall) and construction (introducing spurious elements not explicitly contained in the original materials). Furthermore, even those inferences that are not obviously or directly attributional may nevertheless have important attributional implications. (p.77-78)

Edwards and Potter defend their thesis more by illustration than by demonstration in the form of argument. The book is largely filled with analysis of discourse taken from courtroom transcripts, media interviews, and newspaper articles. They show that, in large part, accounts of events (through recall) are based on the way in which the subject originally attributed causal responsibility (in the encoding process), and that people will fill in details based on the overall causal analysis.

In our terms then, people organize/ encode/ translate the information into and by the appropriate causal schema, and the resulting image is cued based on the trace of that analysis. This is opposed to a version of memory where the unorganized details are remembered and the causal analysis is based on those. Rather, the causal analysis provides the organizing structure for the details (in the same way that the campus locations served as the organizing structure in

Bower's experiment).

Because of the biased nature of Edwards and Potter's samples their argument that all memories are organized according to causal relations is perhaps too general. We should expect the element of responsibility to be prevalent in the legal and political matters they focused on. We can easily imagine memories which do not rely on causal analysis; for example, the recall of the contents of my refrigerator. However, many such memories still rely on other types of schematic organization. Spacial organization will be important in recalling the contents of my fridge ("What is on each shelf?"), and perhaps temporal as well ("When did I last go shopping?").

More generally then, the construction of memories upon recall are based on the schemata which were used in the translation of the original perceptual impulses. The information is organized into various structures, and later recalled based on that organization. Neisser (1967) has argued that this is why Bartlett (1932) observed the tendency to normalize stories upon serial reproduction.

Thus whether one uses perfectly arbitrary relations to commit material to memory (as in a peg-word mnemonic), or whether one uses causal relations (as argued by Edwards and Potter), or whether the information is organized according to some other meaningful schemata (temporal or spacial, and even rhythmic or rhyming groupings), the process is the same. Effectively, the schemata serve as a kind of filing system into which information is pigeonholed.

> The most important advice offered by the many practitioners of "memory improvement" systems is to develop detailed and articulate schemata into which new

material can be fitted. (Neisser 1967, p.288)

The analogy with a filing system, however, is too static; the schemata are constantly altered with each new use. Further, schemata are sometimes formed/ learned by repeated use; other times they are inherent to the organism, spacial and temporal structures are examples. Both Bartlett and Neisser have argued that the schemata are active, and that they form an integral part of the memory itself. It is not as though memory 'traces' are filed away into pre-existing structures which never change. But the importance of the fact that these structures are pre-existing should not be underestimated. Whether they are pre-existing at birth or after years of training, the schemata play an important role in translation. Though the process works both ways, more often than not, the data are distorted to fit the schema rather than the schema being changed to assimilate the data. This phenomena is especially prevalent when the schema is inherited or well used. If we see information gathering processes as hypothesis testing, as suggested, then the phenomena is belief persistence. Our hypothesis/ schemata/ expectancies do not change with the introduction of new evidence nearly as much as normative standards dictate.

Neisser's remarks raises the question of the arbitrariness of our understanding. Partly, it is the function of our knowledge structures to efficiently store data. Now that function may be served equally well by rhyming mnemonics and acronyms as by spacial, temporal, and causal structures. Since no one believes that the structure resulting from a collection of data under a mnemonic corresponds to anything like a natural kind, and since that correspondence is not required for adequate functioning, this provides some ground for scepticism about space, time and causation. The idea is that space, time and causation can serve their psychological function (as structures into which data is stored) without corresponding to anything 'real', just as a mnemonic does. (Such arguments will be examined in detail in sections 6 and 7.) There is a partial way out. Space and time are observable: we have tape-measures and clocks. As such, the data may already be spacio-temporal, and thus there is a disanalogy between these and the arbitrary mnemonics. However, this still leaves causation and other unobservables in a bad spot.

Let us return for a moment to our main question, the function of consciousness. We have seen that organization is a necessary condition for conscious awareness; imagery never consists of unorganized information (see mechanism 2.3). In this section it has been argued that material or information which is schematically organized is more efficiently stored and more easily recalled. Thus we can deduce that information which is present to consciousness will be better remembered. Is memory then the sole function of awareness? Yes and no. No in the sense that awareness, and the conscious system in general, is important for many things beyond recall (see the next two sections). Yes, because the functions to be discussed rely on schematically organized stored information, *i.e.* memories.

Section 4: Behaviour Determination

Presently, we turn to three mechanisms which propose ways in which schematically organized imagery can effect behaviour.

Thus far, we have discussed only the input to the conscious system (the three kinds of non-conscious impulses: perceptual, anticipatory/ suggestions for action, and general), and what that system does with these (translation). We have seen that verbal imagery is, at least in large part, after the fact rationalization. Further, we have seen how these are effective mnemonic devices: that the imagery serves an important function in storage, encoding and recall.

The role of the conscious system in the determination of behaviour relies on mechanisms other than translation and conditioning. In a sense, the three mechanisms which follow can be considered a description of the output of the conscious system. This in turn serves as a kind of input back to the non-conscious. As we shall see however (mechanism 4.3), the communication is not so straightforward.

Mechanism 4.1: inhibition

Recall that it is the schematic organization of the imagistic which separates it from the non-conscious. Recall further that this organization enables the learning of information which cannot be conditioned. Our concern here is with the long term consequences of specific actions. The extended temporal interval between the action itself and such consequences prohibits non-conscious learning. Indeed, we know the consequences of many actions which are never performed at all.

As such, the translation of anticipatory impulses will contain information, not only of the motivations for the anticipated action, but also of the likely consequences. When these consequences are sufficiently aversive, the action will be inhibited by the conscious system. Since the knowledge of long term consequences requires the conscious system, the advantage thereof is apparent. Presumably, expected consequences are desired or aversive based on the inference of our goals from overt behaviour (mechanism 2.3), and can be learned (mechanisms 2.1 and 2.2).

In effect, the mechanism of inhibition serves as a kind of multi-layer filter or screen. Some impulses are blocked before they receive translation (these do not pass one of the first screens), whereas others are inhibited only after the translation is complete and the perceived consequences are determined to be aversive (these are inhibited by one of the later screens). Thus the configuration of the filters will determine which impulses become conscious, and which actions are performed. This configuration is not constant, *i.e.* it will change depending on such things as circumstances (see mechanism 4.3).

Mechanism 4.2: excitation

In its simplest form, this mechanism will add renewed vigour to a non-consciously proposed action. In these cases, the action is only weakly conditioned (non-consciously) to the eliciting circumstances, but once the anticipatory impulse is translated and the consequences recognized/ attributed, the action may be seen to be perfectly appropriate. This might involve recognizing unexpected but desired consequences ("Killing two birds with one stone"). This finds a natural explanation by the trial and error or hypothesis testing view. If a cat is (relatively) randomly attempting various behaviours in order to, say, escape from a cage, it does not expect that licking its paws will yield the reinforcement it is searching for. When the 'random' behaviour is reinforced, thus becoming 'intelligent', excitation is at work.

Further, the translated impulse is not the same as the original non-conscious impulse. That is, it is not as though the impulse simply passes through the series of filters. Rather, it undergoes changes as it travels, the least of which change is translation into imagistic/ schematic form.

If the actual consequences of an impulse which is translated and passed proves to be rewarding, either as predicted or surprisingly, then the translation of the impulse is strengthened. That is, the association between the non-conscious impulse and the specific image becomes stronger; it is almost like the trail which becomes wider as it is trodden. In terms of the filter analogy, we might say that the configuration of the screens are altered to allow the passage of that impulse given certain circumstances. Indeed, as discussed in mechanism 4.3, the configuration of the filters will make it more probable that the appropriate impulse is outputted by the non-conscious system. This then results in what will have the same effect as increased strength of response (the non-conscious response to the particular circumstances).

In a sense then, both non-conscious conditioning and conscious learning can have the same effect on the behaviour of the organism in specific circumstances. Thus whether we learn by conditioning that pumping failed brakes will help stop a moving car, or whether we learn it verbally in driving school (mechanism 2.1), we might still perform the same action given the same circumstances (failed brakes and a moving car).

Importantly, this is another place where we can see crossover between the conscious and non-conscious systems. The same pattern of behaviour can be learned by mechanisms of either. Such considerations show that the two systems are not entirely distinct. For example, some pattern which is learned consciously and originally required awareness can be executed without awareness once the pathways are well established. To take another driving example, one's first experience might be accompanied by such verbal imagery as "OK, depress the clutch, a little gas... slowly release the clutch... accelerate...", while an experienced driver might be able to perform the same actions non-consciously while listening to the radio, talking on the phone, calculating her E.T.A. and chewing gum. Conversely, some behaviour which is originally conditioned non-consciously might later become conscious once a pattern in the flow of non-conscious impulses is recognized and explained (attributed). This is the progression seen in the gambling task (mentioned in distinction 1.2) (Bechara *et. al.* 1997). Presumably, once the subject becomes aware of what he is doing, the pathway between the circumstances and the behaviour is strengthened or reinforced. This strengthening is what is meant here by excitation.

In some cases then, this mechanism is really the same as mechanism 2.2. It is the hypothesizing of an explanation for a perceived pattern (either in the environment or in our own behaviour) which may contribute to establishing that pattern more concretely.

Mechanism 4.3: feedback control/ expectancy

Effectively then, we have discussed two mechanisms by which the conscious system can effect non-conscious processing. The first allows for the inhibition of impulses determined to be inappropriate, and the second enables the smooth progression from situation to action. However, this leaves open the case where the situation is novel; here, there is no predetermined pathway (established by mechanism 4.2) and the schemata of the conscious system might be necessary for determining a course. Mechanism 4.1 might be insufficient by itself when all of the non-conscious impulses (suggestions) are inhibited. The conscious must serve as a guide to the non-conscious: a simple "yes" or "no" (passed or inhibited) will leave the latter with nothing but trial and error as a way to generate a passable impulse. That would not constitute 'intelligent' behaviour.

The way the system has been construed so far, there is no place for the conscious system to generate impulses on its own. That is, action must first be suggested by the non-conscious, then pass through the filters of the conscious system before it can be executed. There is no room for the conscious system to elect some action based on perceptual impulses and schematically organized information. Only for it to determine which impulses are passable. Though such a mechanism is not *a priori* impossible, neither is it necessary or desired. Such a mechanism might revert to intentional or conscious states as direct causes of behaviour. Here, the conscious system is capable (so far at least) only of translation and inhibition. Though it has not been argued explicitly or convincingly, it has been accepted that an impulse to action must originate in the non-conscious system. So how then, can the conscious system effect the impulses from the non-conscious besides simply translating and inhibiting? That is, how can it inform the latter of what is appropriate? For example, suppose I receive an invitation to go to the beach. The non-conscious impulse will be translated into the schematic image of an afternoon at the beach, which will probably include some prediction of what will happen and whether it is desirable or aversive. Suppose too, that the translation contains the information that my ex-girlfriend (whom I do not wish to see) will probably be at the beach today. This will be sufficient to inhibit the suggested action (going to the beach). But what of the possibility of going to a different beach? On the theory in development here, that will only occur to the conscious system if it is first suggested by the non-conscious. There is a conflict here between the desire to go to the beach and the aversion to seeing my ex-girlfriend. How can the non-conscious system 'know' that although the impulse to go to the beach was inhibited, the impulse to go to a different beach will not be? How can it know why the impulse was inhibited, and thus which impulse will be passed? By the mechanism of feedback control. The conscious system controls the output of the non-conscious system. That is, the conscious system controls its input.

A classic feedback control system will have the following properties: first, the purpose of the system is to maintain a controlled variable equal to a reference value despite external disturbances, second, the system will have a feedback path, and third, the system includes a sensing element and a comparator (Mayr 1969).

Let us examine a simple example: a thermostat. The controlled variable is room temperature, the reference value is the setting. The thermostat senses the former and compares it with the reference setting, its output is a function of the difference. Feedback is present because the input to the system is effected by its own output. Now, we might say that the function of the thermostat is to offset external disturbances; it must compensate for heat loss. But it does not measure heat loss. This would involve measuring such things as outdoor temperature, wind velocity, insulation, the colour of the roof shingles, *etc.* It cannot control a variable which it cannot measure. Thus we say that the thermostat does not regulate its output, *i.e.*, it does not regulate the heat produced by the furnace. Rather, it regulates its input: the sensed value of the room temperature. "Having set the thermostat, one can predict the indoor temperature but not the fuel bill" (Hershberger 1990, p.57).

This is the crucial feature of a control system: it controls its input. Its function is to maintain that input within some interval specified by the reference signal. The system compensates for external disturbances by offsetting their effect on input through action. In the biological context, the reference signal is construed as the goal state, and the input as afferent stimuli. Thus the organism acts to maintain sensory input within established limits.

This construal of goals as states of input is not new.

William James (1890), although not familiar with the mechanism of control, wrote that voluntary action is that which is directed at some goal, and that the goal state consists of sensory consequences and not muscular innervation. That is, voluntary action is directed at some state of input, not output. Some of the early behaviourists, such as Watson and Hull, made the mistake of construing learning as the associating of sensory conditions with specific motor engrams. Tolman quickly pointed out that this could not be right, else a subject who learned to move her finger upward from an electrode would perform the same movement were her hand turned over, thus driving her finger into the electrode (Leahey 1991). Around the same time, Karl Dunker (1935) also argued that the "reflex doctrine" was in error on this issue. This matter will arise again in section 6.

James (1890) distinguished intentional or intelligent action from the mechanical by the fact that with the former, it is the end which is fixed while the path varies, whereas with the latter only the path is fixed. Construing ends or goals as desired states of sensory input will suffice for a wide range of such intelligent behaviour. In the next section, when we turn to problem solving, the notion of goal states will be expanded to include any input to the conscious system which satisfies the constraints specified by the problem.

The remainder of this section deals with global properties of control systems, and applications to relatively simple behaviour. The next section will deal with an adaptation of this mechanism to more complex conscious reasoning.

Consider an example of a control process (Hershberger (1990). When driving a car on the highway, we maintain the focus of expansion in the visual field in the centre of our chosen lane. The input is then the location of focus of expansion, and the reference signal corresponds to the center of our lane. We control the input by mechanically moving the steering wheel. Now as drivers, we must compensate for such potential disturbances as curves, the slope of the road, and cross winds. This is not achieved by measuring each and employing a complex formula to calculate the appropriate position of the steering wheel. Rather, all such potential disturbances are automatically offset when the location of the focus of expansion is controlled.

Now consider how one might learn to be such an expert driver. Assume that the novice driver already understands that his job is to reduce the difference between the location of his focus of expansion and the center of his lane, *i.e.* to maintain the visual error signal as close to zero as possible, and that he knows that he can direct the car by turning the steering wheel. Now an increase in the error signal will result in positive feedback and aversion, while a decrease will yield reinforcing negative feedback. Thus the stage is set for operant conditioning to take place. The driver adopts the previously discussed tactic of trial and error. When an error signal appears, he randomly adjusts the position of the steering wheel. If positive feedback results, he will change strategies; if negative feedback results, learning occurs. He might learn, for example, to turn the wheel to the right if the focus of expansion drifts to the left. Hershberger (1990) has maintained that all operant conditioning can be explained similarly by control processes. Further, Powers (1978) has shown how the mathematics of control system engineering predicts some of the classical results of conditioning experiments concerning the frequency of the operant under various schedules, and how the classical stimulus-response analyses of behaviour can be considered a special case of control (zero loop gain: no feedback).

Hershberger goes on to demonstrate how Pavlovian conditioning is also a control process. In effect, the organism will create its own disturbances to preempt an anticipated external disturbance. Whereas operant conditioning teaches us to compensate (feedback), classical conditioning teaches us to anticipate (feedforward). Each time our (now competent) driver passes a truck moving in the opposite direction a pressure wave (unconditioned stimulus) pushes his car toward the shoulder. For the first few, he will compensate for the disturbance by steering the car back toward the center (unconditioned response). But soon he learns to associate the sight of an oncoming truck (conditioned stimulus) with the unconditioned stimulus. The laws of classical conditioning tell us that he will learn to steer into the truck (conditioned response) as it passes. Thus rather than waiting for the disturbance and then compensating, he maintains his error signal at zero by anticipating the unconditioned stimulus and generating his own internal disturbance (conditioned response) so that the net disturbance is zero.

As presented, the mechanism of control can cover alot of ground. If it accounts for conditioning, then it probably accounts for a great deal of our non-conscious behaviour. That is, we have the conscious controlling the non-conscious, and the non-conscious controlling the sensed environment. Powers (1973a, 1973b) has proposed a hierarchy of control systems which he argues can account for all organic behaviour. I intend to loosen up the tight, engineering inspired closed-loop negative feedback control system analogy so as to make it accessible to introspection, and adaptable to problem solving.

I will generally speak as if the conscious system controls the non-conscious. (Though practical for explication, this is not entirely correct. Rather, we should speak of thousands of small feedback loops, and loops within those, with higher levels of processing controlling the lower levels and consciousness arising according to the strength of the imagery.) The reference setting is then some desired state of input to the conscious system, *i.e.* output of the non-conscious. This can be construed as an image (either loose or vivid) such as "I wish I had a cigarette", or analogically as the configuration of the filters. The loop is closed (feedback is maintained) because the output of the conscious system in turn effects its input. But the crucial feature of this construal of feedback control is the expectancy.

Generally, a control system will have an output signal which in turn effects its input. But this is not required here. Though it is possible, I have something else in mind. Whether we describe the output as a distinct signal sent back to the non-conscious system, or as simply the setting of the input specifications (the reference signal), as we will describe it here, the two systems can be functionally identical. However, if the output of the conscious were simply the re-configuring of the filters, then we would be stuck with the same problem with which this subsection began. We would still require another mechanism so that the non-conscious was not relegated to trial and error tactics.

I propose that the specific configuration of the filters, the value of the reference setting, creates a kind of expectancy. This means only that the reference setting itself will make it more probable that the non-conscious system will output an impulse which reduces error, or will lead to the goal state, or is passable and desired. Construed as an output signal from the conscious system, the content is a "Gimme something like this". But if we prefer, we can consider the conscious system as 'drawing' the appropriate impulse from the non-conscious. The reference setting acts as a kind of 'active keyhole', which pulls for the impulse with the appropriate fit. Either way, the necessary function is fulfilled: 'the outputting of the appropriate non-conscious impulse becomes more probable'. This is only a slight warping of the engineering control systems, since there too we might construe the reference signal as pulling the input signal into line with it. Thus we conceive goal states or desires as an active component in their own satisfaction.

Continuing with the beach example, we can now see why the non-conscious system is not blind when its suggestion is inhibited. The conscious system translates the impulse as before, and again the image will contain the information that while going to the beach is desired, seeing my ex-girlfriend is aversive. Thus the filters are organized into a configuration which will draw an impulse which will be translated into an image which includes going to the beach but excludes seeing my ex-girlfriend. Going to a different beach is an obvious possibility. (This is an example of analysis of conflict which will be explained more fully in the next section.)

We might interpret an organism as a locus of reference signals. It might have settings for body temperature, sleep, hunger, caffeine, material goods, social interaction, respect, dominance, sex, *etc.* When the input level for any of these falls significantly far from the reference value, the organism will seek to alter the former through action. Thus if the body temperature is too low, goosebumps or shivering might result, if energy levels are low, the organism might seek out food, if respect is lacking, the organism might buy a sports car. Conversely, the level of direct exposure to the sun might be too high and shade or clothing is sought. The reference value represents the level of satiety.

Now some people might not have an upper limit on their satiety level for, say material wealth. Nonetheless, they must pause their efforts to increase that input value to eat, for example. Thus we see a hierarchy of desires. The difference between the input and some reference value will not always be dealt with immediately. Often there is a more important desire which must be satisfied first, or we might allow an error signal to persist so that some other long-term goal may be satisfied. Other times, certain reference values are created only subordinate to some other goal: "Find a gas-station so that I can fill my bicycle tire so that I can get home soon so that I can get to sleep early so that I'll be fresh tomorrow so that I'll have a better chance at the job interview so that...". Further, a problem can set a novel reference setting.

It should also be noted that not all reference settings will draw an impulse to the same degree, or even at all. A reference setting may be considered a desire according to the degree that it does draw for its satisfaction. Some actions or impulses may be passable but not drawn for, non-aversive but not desired. Thus this mechanism provides an analysis of the term 'desire' which closely mirrors our common usage. There is also a sense in which the strength of the anticipation or draw is dependant on strength of imagery. This connection will be discussed in the next section. This same kind of expectancy driven process, this pulling and resonance, has been used to explain placebo effects (Einhorn and Hogarth 1978), as well as the constructive aspects of memory recall, resonance, translation/ attribution, and confabulation. It has also been used to show how self-fulfilling prophesies are possible, whereby "reality-testing becomes reality-construction" (Snyder and Swann 1978). It is the intentionality, the going for, searching out, and fulling forth: the 'making it more probable that'.

Though the discussion of behaviour determination has been brief, we now pass on to a discussion of problem solving. This is because much of what can be considered decision making is really just a special case of problem solving (the beach example is one such case).

Section 5: Problem Solving

I would like to adopt a framework for problem solving used by Dunker (1935), and more recently by Newell, Shaw and Simon (1963). Though their respective structures are not identical, they share some features which will prove important to our analysis. Mechanism 4.3 will be applied to this framework once it is explained. The importance of strength of imagery, in its various forms, to problem solving will also be illustrated.

Definition 5.1: the family tree or problem maze

The mentioned framework can be graphically represented by a family tree. The problem itself is at the top of the tree, and each possible (tentative or proposed) solution constitutes a branch. Each branch may be further subdivided into more specific tentative solutions, one or more of which may be an actual solution. The groupings are based on the functional value of a possible solution method. That is, the principle by which a proposal might lead to a solution (the answer to the question "Why is this a solution?") (Dunker 1935).

Very similarly, Newell Simon and Shaw (1963) have proposed a framework based on an analogy with a maze. Now the maze is not physical, but abstract. It consists of a set of pathways through nodes or choice points, one or more of which will be a solution pathway. Thus problem solving consists in finding a subset of pathways (or just one unique pathway) which satisfies certain conditions. Often, the problem itself will define the entire set of possible pathways (the problem space), as well as the criteria for a solution. This is the same framework used in the design of the Logic Theorist and the General Problem Solver programs.

In our terms, the schematic organization of the conscious system can be seen as constituting a large maze or tree. The posing of a problem will set the initial conditions (the givens) as well as the solution criteria, thus defining a narrower, more delineated tree. The problem is solved when the correct pathway is identified as constituting the solution. Thus it is the same schematic network discussed in the realm of memory, only reparsed (analogically) as a physical structure.

Consider, by way of example, Dunker's most experimentally studied problem:

> Given a human being with an inoperable stomach tumour, and rays which destroy organic tissue at sufficient intensity, by what procedure can one free him of the tumour by these rays and at the same time avoid destroying the healthy tissue which surrounds it? (p.1)

The tree for this problem consists of the problem itself, and that part of the schematic network which is relevant. Plausible first branches (groupings by functional value or nodes) might be: first, "Avoid contact between the rays and healthy tissue", second, "Desensitize the healthy tissue", and third, "Lower the intensity of the rays on their way through the healthy tissue". These are also reformulations of the problem (Dunker's terms), or sub-goals (Newell, Simon and Shaw's terms). Each of these might be further pursued until either it becomes obvious that the branch will not yield a solution, or a solution is found. For example, the subject might try to achieve the first subgoal by suggesting using the esophagus as a free path to the stomach, or by suggesting the insertion of a cannula (these are further branches from the first). Once these are seen as fruitless, the subject might further pursue the second or third branch, until he hits on the preferred solution: using a lens to concentrate diffuse rays at the tumour, a subbranch of the third.

Now the problem itself can be seen as a question, or simply phrased (imagistically) as a question. Further, each choice point can be considered a reformulation of that question, and each will narrow the search parameters, or cut off branches. If the solution pathway contains the specific choice point which is the reformulation, we are making progress, if not, we are chasing a wild goose (this can lead to fixedness). Importantly though, the solution itself is just a reformulation of the problem, so sharp that nothing more is needed. Each reformulation/ branch consists in the variation of one or more of the elements of the problem itself.

> The final form of a solution is typically attained by way of mediating phases of the process, of which each one, in retrospect, possesses the character of a solution, and, in prospect, that of a problem. (Dunker 1935, p.9)

These "mediating phases" are the reformulations. Thus, on this framework, problem solving consists in successive reformulations of the problem, each of which serves to narrow the problem space, until it is sufficiently narrow so as to specify exactly one pathway (or one node) which is the solution. As will be illustrated, the narrowing of the problem space by reformulation serves to aid solution by focusing our limited attentional resources (by increasing the strength of imagery) on a more precise question. A reformulation of the problem is akin to a hint, insofar as the actual solution lies on that branch.

Mechanism 5.1/ 4.3: resonance

A question will serve as a kind of retrieval cue for its answer. It creates the expectancy and anticipation discussed in mechanism 4.3 (feedback control). Often, the question will specify a unique answer ("What is the date today?"). However, many questions cannot be answered immediately by pulling forth the relevant solution. This will be the case if, for instance, the answer is not stored as the answer to the specific question or an analogous question. This is the case in problem solving: the question does not lead directly to the solution, or through a series of pre-defined steps (such as in a long division task). Indeed, this is what constitutes a problem. This is why our metacognitions concerning whether we will be able to solve a certain problem are considerably less accurate (indeed almost non-existent) for problems requiring 'insight' than for algorithmic or strict memory retrieval questions (Metcalfe 1986). That is, a question is a problem only when the steps necessary for its solution are not immediately recognized. We might say that the solution is not directly accessible from the problem. Indeed, this is where we will need to reformulate or restructure the question into a form from which the solution is accessible.

The essence of solutions by resonance is the pulling of the goal state or solution through a pathway which ends (begins) with the problem. This is why we are not relegated to a trial and error method of problem solving, in the same way that we need not rely on trial and error to locate an action which will not be inhibited; resonance is really just an application of mechanism 4.3. Solutions through resonance form the essence of all kinds of problem solving according to Dunker, and fortunately, it is conformable to control theory.

The problem can be expressed by ?Rb, where R is the relevant relation and b is the known image. For example, ?Rb might be interpreted: "Something tending to cause the effect of warmth", or "Something to use in place of a hammer". The statement ?Rb is a kind of question. It should be evident that its content is exactly what we should expect the content of the reference signal to be (*i.e.*, "Gimme something like this", or "Gimme something which tends to cause..."). In using this method, the question serves as the model of search; it is the schematic network, or sometimes the perceptual field, which constitutes the region of search. The solution is found in virtue of the similarity between the property demanded (Rb) and a property of that which is sought.

> The problem is: ?Rb; aRb exists in the thinker's experience; by reason of the partial correspondence with ?Rb, aRb and therefore a are aroused. Thus this finding of the solution takes place ultimately through a kind of "excitation by equality"... or, better, of resonance. (Dunker p.19)

As stated, this setting of the reference value will increase the probability of something with the appropriate fit being impulsed from the non-conscious system. We have spoken of this as a pulling from the conscious for the appropriate impulse. In the case of problem solving, the reference setting is the question or problem, and it pulls for its own solution. This provides the intelligent direction for an otherwise trial and error procedure. Dunker's preconceptions about the working of the mind lead him to construe the problem as a model of search, and thus problem solving as a process of memory recall. My own preconceptions have led me to construe the problem as designating a reference signal (?Rb) which pulls forth the solution from the region of search. There is really not much difference, save for the fact that my construal is more conformable to control theory. (A plausible connection between control theory and memory recall will be discussed in the next section.)

Strength of imagery revisited

Now, it is a consequence of our limited resources that not every question will be able to pull out its answer directly or immediately. This is the function of the reformulations. They represent intermediate steps from which the answer (that which is sought: the "?") may be pulled. Often, when the initial question is not strong enough to pull the solution, a reformulation will be. The reformulation is one way to increase the strength of imagery, which in turn makes the solution more probable. The reformulation narrows the problem space such that it might pull further, but down a more specific pathway, thus making more efficient use our limited attentional resources. This can be helpful or detrimental. If the reformulation is on a solution path, then it will probably aid in the solution process, if not, it might contribute to fixedness.

That a solution can be accessible from one question but not from another when these questions contain the same information can be demonstrated in the following example.

> If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?

This question was first posed by Casscells et al. (1978) as a demonstration of base-rate neglect: 45% of fourth year Harvard medical students answered 95%. The question, and numerous variations of it, was also employed by Cosmides and Tooby (1996) in order to show that by reformulations of the problem into frequentist terms could eradicate the base-rate neglect and improve Bayesian reasoning. It is used here to exemplify the point that the subsequent reformulations improve results because they allowed for a stronger image of the question, or a reformulation from which the solution was more accessible.

Imagine the reasoning involved in solving this problem of someone familiar with probability theory but unfamiliar with Bayes' theorem. (The problem is purely algorithmic for someone who employs Bayes' theorem.) Perhaps the first image following perception of the problem would be the verbal image "51 out of 1000 people will test positive for the disease". Next, "Only one of these people will actually have the disease", followed by "One out of fifty-one people who test positive actually have the disease", and "Two out of 102 people who test positive will actually have the disease", and finally, "Approximately 2% of those who test positive actually have the disease".

Compare this with one of the reformulations employed by Cosmides and Tooby:

> 1 out of every 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive (i.e. the "true positive" rate is 100%). But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease (i.e. the "false positive" rate is 5%).
>
> Imagine that we have assembled a random sample of 1000 Americans. They were selected by lottery. Those who conducted the lottery had no information about the health status of any of these people.
>
> Given the information above: on average, how many people who test positive for the disease will actually have the disease? ___ out of ___

In the original version, 12% of subjects answered correctly. In this revised version, 56% of subjects answered correctly. Further, the final paragraph was replaced with:

> Given the information above: on average,
>
> (1) How many out of 1000 people will have the disease?
>
> (2) How many of the 1000 people will have the disease AND test positive for it?
>
> (3) How many of the 1000 people will be healthy AND test positive for the disease?

In this version, 76% of subjects answered correctly.

There is no new information in the reformulation presented here. That is, there is nothing which could not have been derived from the original question. There was, however, a more than six fold increase in the proportion of subjects who answered correctly. It is argued that this is due to the different, more explicit structure which yields a stronger image of the question. Further, the account (a few paragraphs back) of the successive images one might have experienced upon solving the original problem was meant to illustrate that the reformulation represents an intermediate step in the process. As such, a subject faced with the first question is required to establish a pathway from the question, through the reformulations to the solution. Whereas the subjects given the reformulated question were required only to make the second part of the trip. "A solution will be the more difficult to find the more work of explication it presupposes" (Dunker p.40).

Now, a reformulation can serve to strengthen imagery. The subjects who were asked to answer the original question must hold something close to the successive reformulations in consciousness in order to solve the problem. They are at a disadvantage; the solution will demand more from their attentional resources. The image resembling the reformulation will necessarily be weaker than the same image will be for the second group who have access to the image in perception. Further, the reformulation is stronger in the sense that it is closer to the solution, or that the solution is more accessible from it. In Dunker's terms, the phase distance is reduced. Thus strength of imagery is increased (both in embeddedness and vividness), and it is argued that it is the increase which yields the higher probability of solution.

Another way in which strength of imagery is helpful in problem solving is demonstrated by Kohler's studies (cited in Vygotsky 1934 and Dunker 1935). He observed, for example, that chimpanzees were capable of learning to use a stick to reach through the bars of a cage to reach a banana, and that they could make a tool, in the form of a sufficiently long stick, by inserting one short stick into an opening in another. Lesser known results of Kohler's work were that the primates would not initially think of employing the tool unless the stick and the fruit were in the visual field simultaneously. Further, they would not realize the possibility of constructing the longer tool if the two short sticks accidentally crossed in their hands, forming an 'X'. This demonstrates again that increased strength of mental imagery (vividness in this case) leads to a higher probability of solution.

The chimpanzee is anticipating/drawing for "Something with which to reach the food". As previously stated, an image is stronger when it is present to immediate perception than when merely imagined. As opposed to the previous case where the question was stronger, here, it is the answer which is stronger. It will be so when the two short sticks are laid out on the ground end-to-end than it will be if laid in an 'X', *i.e.* when the configuration in perception more closely resembles that which is sought.

In problem solving using this mechanism, each system serves a function. It should now be clear how the strength of imagery will help, but it is still the duty of the non-conscious system to impulse the solution. That is, the reasoning does not take place consciously, only the questions and reformulations. Careful introspection will reveal this fact. If the problem is understood correctly, then the solution pathway will be open (not inhibited), but with insight problems, the pull will not be strong. The non-conscious must continually output impulses until it finds one that fits. These inhibited impulses need not be translated into imagery. Further, even once an impulse is passed, the solution must still be verified. This is why we often find, in experience, that the solution we seek occurs spontaneously to awareness.

Section 6: Confabulation and Gap-Filling

In a now classic paper, Berlyne defined confabulation as "a falsification of memory occurring in clear consciousness in association with an organically derived amnesia" (1972). Since then, there have been a multitude of studies associating confabulation with various disorders and pathologies. Stuss *et al.* (1978) have argued that amnesia is not a sufficient condition for confabulation, frontal dysfunction must also be present. Specifically, these authors attribute the tendencies to the failure of inhibitory mechanisms and self-monitoring capacities, the misuse of environmental cues, perseveration, impulsiveness, and lack of concern about incorrect performance. Shapiro *et al.* (1981) confirm the correlation of confabulation with all but the last two listed deficits. Fischer *et al.* (1995) agree, adding that the severity of the confabulation is proportional to the degree of executive system deficits and frontal lesioning.

These three papers all deny that memory gaps produced by amnesia are sufficient to produce confabulation. However, they are largely concerned with what Berlyne called "fantastic" confabulation. This type is characterized by grandiose and sometimes bizarre spontaneous verbalizations which often change from day to day, though sometimes are stable and repeated. In some cases, the confabulation takes the form of a confused outpouring of irrelevant associations. For example one of Stuss' subjects at the Boston VA hospital initially reported that he was in Finland.

> While sitting in the corridor of the ward in his pajamas as the dinner carts were wheeled in, he

> explained that he was in a navy kitchen. That afternoon, after the trays had been removed, he described an almost identical external environment as a "storage area for hallways"... when asked to accompany a physician, he stated that he was waiting for "Billy and the boys to come help fix the pipes" (p.1167).

In contrast to fantastic confabulation is "momentary" or "provoked" confabulation. This second type is generally produced as an answer to questioning. It is usually more modest and plausible, experimenters often need to verify the subject's reports in order to determine whether they were confabulatory. Also, this is the same type of confabulation observed in the recall of events or information. It involves the introduction of erroneous and fabricated material. A favourite example comes from Gazzaniga and Ledoux (1978). Instructions are given to the right hemisphere of a split brained individual, for example, "walk to the door". Once the instructions are carried out, the subject is asked "why did you walk to the door?". At this point, it is the left hemisphere which responds. Now the left hemisphere, though aware of the experimental setup, was not privy to instructions and so confabulates a plausible response: "I was thirsty, I was gonna go into the hall for a drink".

The distinction between fantastic and provoked confabulation has been called into question by the fact that some fantastic confabulators become mere provoked confabulators as they recover to 'non-confabulators' (Weinstein, Kahn, and Malitz 1956, Shapiro *et al.* 1981). Others, however, have argued that
the two types represent different neurological deficits.

Joseph (1986) maintains that fantastic or spontaneous confabulation is due to frontal damage resulting in the failure to inhibit inappropriate verbalizations and speech release due to the flooding of the language areas with irrelevant and grandiose associations, while provoked confabulation is a result of gap-filling where information is lost due to disconnection, amnesia, or lesioning.

"Confabulation" seems to be something of an umbrella term for various modes of the production of verbal imagery and speech. There are, however, a few elements which the various forms have in common. First is the fact that confabulated material is generally false. It is also fabricated or constructed. We should also distinguish confabulation from lying; confabulators do not intend to deceive, though they do succeed in deceiving even themselves. That is, the confabulations are not offered as a guess or a suggestion, but as firm and established fact. Lastly, confabulation, very broadly construed, is not restricted to the verbal modality.

Further, I would argue that many cases of provoked confabulation can be explained by the gap-filling aspect of mechanism 2.3 (translation). Recall that two of the paradigm cases of translation discussed in section 2 were the consumer survey of Nisbett and Ross (1977) and the construction of memories upon recall made explicit by Bartlett (1932) and others. The example of provoked confabulation in this section was chosen specifically for its striking resemblance to the former. In both cases, the subject invents a reason for the behaviour based only on the observation of the behaviour itself along with theories concerning which motivations are the most plausible for such actions.

Kopelman (1987) has argued that provoked confabulation is a normal response to a less than perfect memory. It is well known that amnesia and confabulation are common to both Alzheimer and Korsakoff patients. Kopelman compared the frequency and nature of confabulation between two groups of such patients and a group of healthy control subjects. He employed stories from the Wechsler Logical Memory test, examining the confabulations upon recall. The amnesia in the experimental groups was controlled for by testing the recall in the controls after an extended interval. Despite the fact that the controls scored better after a week than either experimental group on immediate recall, the frequency of confabulation was comparable across all groups (almost equal in the control and Korsakoff groups, and slightly lower in the Alzheimer's group), and the nature of the confabulations were identical. Kopelman himself remarked on the similarity between the constructive aspects of recall found in all three groups, and Bartlett's findings of intrusions and distortions.

The point is simple: though fantastic or spontaneous confabulation may occur only in association with frontal lobe dysfunction, provoked confabulation occurs both in amnesiacs and healthy people. Whereas amnesia is not a sufficient condition for spontaneous confabulation, it is sufficient, though not even necessary, for the provoked type; we are all confabulators.

There is an interesting epistemic question in all this. Can we distinguish between what is confabulated and what is not. If we could somehow divide the contents of the schematic network, then we could assess the epistemic grounding for our various judgements. They would be better justified insofar as they contain less confabulation. Now we cannot always rely on an experimenter with privileged knowledge of the situation to tell us when we have erred, as in the consumer survey or the split brain patient or the recall of recorded linguistic information. Even if we could, this would not suffice since what is confabulated is not co-extensive with what is false. Neither can we distinguish introspectively, since it is a feature of confabulation that it is believed by the subject to be true. Nor can we rely on general opinion, since that would invariable lead to epistemic relativism. Let us look more closely at gap-filling.

Now, both Kopelman and Joseph agree that provoked confabulation (henceforth "confabulation") is produced by a gap-filling process. But the interesting aspect of any such process is that it requires something between which there are gaps to be filled. Let us call this something the "data points". If we can distinguish the data points from the confabulation in a way which will allow us to classify the contents of imagery, then we will have located an epistemic foundation. This foundation will inevitably be far from infallible, but it will also be better grounded than what it excludes. That is, we will be more justified in believing that our data points correspond than in believing that the confabulation does.

Fear not, I do not intend to invoke any kind of pretheoretical knowledge, for that seems impossible. Nor will I attempt any receptivity-spontaneity distinction (*e.g.* McDowell 1994, Tye 1997), because in the end that relies on the notion of a will which was discarded sections ago.

Rather, my plan is to demonstrate what will be sufficient as data points to account for the majority of human behaviour. Because of what has been said before, this will amount to showing what is sufficient for each of the mechanisms I have proposed. If it can be shown that some proper subset of the contents of imagery is sufficient for the adequate functioning of the organism, then anything beyond that can be considered confabulated. It is argued that the proper association of afferent stimuli is sufficient, and therefore can serve adequately as the data points.

Let us begin with conditioning. Now conditioning is of two sorts: operant and classical. I propose that the association of afferent stimuli is sufficient to account for the knowledge acquired in either case. Consider operant conditioning. It is often thought that what is learned here is a stimulus-response association (as discussed in section 5). However, it seems that the only major S-R theorists were Watson and Hull (Leahey 1991). Following Skinner and Tolman (and James and others), the operant is best viewed not in terms of motor engrams but in terms of a certain state of sensory feedback. Otherwise, a rat who learned on the first few trials to take three steps forward and press the bar with its right paw when the light goes on in the Skinner box (*i.e.*, a certain specific motor engram) would not be able to adapt to a situation where its starting position was different or its right paw immobilized. Thus what is learned in this case is an association of the circumstances (the Skinner box with the light on) with the rewarding afferent state (the bar being pressed), both construed in terms of sensory stimuli.

Let us move on to classical Pavlovian conditioning.

Now these cases rely on an existent association between an unconditioned stimulus (UCS) and an unconditioned response (UCR). (Here again, the response is best construed as a specific afferent state.) When a new stimulus is consistently paired with the UCS, it becomes a conditioned stimulus (CS) and it will elicit a conditioned response (CR) resembling the UCR. The CR and the UCR are in effect the same behaviour, the learned association being between the UCS and the CS, both of which are stimuli by definition.

Now it might seem that since conditioning aims at the association of stimuli, there is no room for confabulation. This is not so. Confabulation is a process of gap-filling; data (afferent stimuli) is moulded to fit expectancies. The organism does not start out with a *tabula rasa* onto which associations are recorded. Rather, it already comes replete with expectancies in the form of genetic predispositions. Recall the discussion of section 1 which describes conditioning as a trial and error or hypothesis testing process. Even in conditioning there is some hint of top-down, conceptually driven learning.

Consider Watson's famous experiments with little Albert. He was able to induce a fear reaction to a rabbit by pairing the appearance of the rabbit with a loud noise. However, when one of Watson's graduate students attempted to induce a similar reaction to a block of wood or a cloth curtain, he was unsuccessful. Similarly, rats will learn on only one trial to avoid a new tasting food when it is followed by gastrointestinal illness. The expectancy is so high in this case that the temporal interval can exceed twelve hours, however, if the food is familiar tasting but a new shape, conditioning will not occur. Conversely, rats will learn to avoid food of a new shape if followed immediately by an electric shock, but they will not avoid food of a new taste if followed by the same. It is also notoriously difficult to teach cats to lick their paws, or dogs to yawn for reinforcement (examples from Nisbett and Ross 1980).

The point is that not all associations are learned with equal ease. All organisms have preconceptions or expectations about the plausibility of various associations. Confabulation is the result of the biased approach to hypothesis generation, and the confirmation bias evident in Skinner's pigeons. (It might be noted that the predispositions will generally serve an organism well if the environment is similar to that in which the predispositions were initially selected. However, our industrialized world might be construed as significantly different from the world of our ancestors, and surely the atomic universe is too far from that to trust our preconceptions about what will be found there. Though we are well advised to trust our intuitions in natural circumstances, there is doubt as to whether scientific theories should more likely be true if they are intuitive.)

Thus the information learned in both types of conditioning is the association of stimuli, and that is what serves as data for this kind of knowledge: afferent states or observable circumstances. Now it is somewhat strange to call what is learned here "knowledge", since it generally occurs non-consciously. However, it is this type of association which constitutes the information which guides all of our behaviour. As noted (section 1), the mechanisms of conditioning cannot account for much of the knowledge we all have, because they rely on, for example, the close temporal proximity of the stimuli to be associated. This is why translation and feedback control were introduced. I maintain, however, that it is exactly the same kind of associations which constitute the data points required for these mechanisms to function.

My claim then, is that the data points of our semantic knowledge consists only in the association or relation of observables. Again, my argument will proceed by a demonstration that the association of stimuli is sufficient data on which to base our judgements and decisions. This in turn is demonstrated by examining the major mechanisms proposed here. The behaviour explained will be divided into two types: everyday behaviour guided mainly by conditioning and aided by translation and feedback control, and the more complex behaviour as exemplified by science. The remainder of this section is devoted to the first type, showing how the mechanisms of translation and feedback control can function adequately on the assumption that stimulus associations are sufficient construals of the data points. The next section will deal with the observation-theory or data-confabulation dichotomy, and the application of this view to the more complex semantic knowledge.

Recall that translation is the mechanism whereby non-conscious impulses become conscious. In the process, the data becomes reorganized and categorized, attributions of causes, internal states, attitudes and the like are inferred. Recall also that translation has two functions: the efficient storage of information and social justification. It is my claim then, that neither of these functions requires anything more than the association of stimuli as data. That is what must be preserved through translation, the rest may be confabulated without detriment.

Suppose one employs a mnemonic device to store a list of unrelated words, as in Bower's (1970) experiment. Then, so long as the words are recalled accurately, the specific mnemonic used is irrelevant. The original list constitutes the data in this case, the mnemonic is the confabulation. Note that the mnemonic need not be something obviously confabulated, as Bower's method of loci was, but may seem genuinely true. I will argue in the next section that the theory of unobservable physical particles is such a case. For now, recall Neisser's (1967) remarks quoted in section 3 to the effect that, as far as memory functions are concerned, what is important is that we have some elaborate schemata for storing data, and that the specific choice of theory is unimportant. Thus, if I am right and the relevant data for our everyday behaviour consists only in the association of observables, then insofar as our semantic theory goes beyond that, it is confabulation.

With regard to social functions consider again the case of Nisbett and Wilson's (1977) consumer survey. Here the causes of the subject's behaviour were inferred from the observation of their own action, the observable elements of the situation, and semantic theory concerning socially appropriate reasons. In this case, the action and the situation were the observable data, while the semantic theory employed in the translation served only the purpose of social justification. Were our society one where the quality of pantyhose was determined by size, then likely most of the subjects would have reported that they chose the largest pair. The point is that to serve the function of social justification, the translation need not correspond to reality, but only to social norms. (It must of course also agree with the observables, in this case the action in the situation.)

Thus translation must preserve the observed, but beyond that, its functions may be served equally well by confabulation as by true semantic theories. This is not to say that inference is never truth preserving, or that what is not observed never corresponds. For example, in the case of the reverse-incentive effect (as discussed in the context of the Festinger and Carlsmith (1959) experiment in section 2), though the attitude toward the tasks was attributed/ translated from the observable data, it was not necessarily false. Suppose we operationally define such an attitude as the subject's report of it, then the translation is trivially true. Or suppose we define it as the subject's willingness to participate again. Then it might turn out that the attributed attitude was the correct one. My point is only that the mechanisms which I've offered to explain human behaviour are consistent with the idea that much of the content of our imagery is confabulated, and that it is sufficient for efficient everyday functioning that the data points consist only in the association of stimuli.

The final mechanism to be examined here is feedback control. The first thing to note is that the data relevant to this mechanism cannot be anything more than the association of stimuli, if what I have claimed is correct, since feedback control introduces no new information into the schematic network.

The first kind of application of feedback control is to simple non-conscious control systems, those which regulate such things as body temperature and hunger. Assuming a relatively constant reference setting, it is the function of these systems to maintain the input value within the limits of the reference interval. Now the input for these simple control systems cannot be anything but afferent stimuli. The fact that it is input which is controlled is best exemplified by such simple systems. There should therefore be no doubt that the required data is just the association of afferent stimuli: sometimes observed circumstances are associated with a specific reference signal (*i.e.* that reference value is triggered by the circumstances), and sometimes a response is associated as the preferred way to satisfy a reference signal. As discussed, circumstances, reference signals, and responses are best construed as afferent stimuli.

The case of practical problem solving poses a little more difficulty. What is required here is that ideas and images be related in the appropriate ways. We require intelligible relations and understanding. Recall that to solve such a problem, we must be able grasp the functional value of the solution, *i.e.* the reason why it is a solution. Now the injection of higher order images and understanding may seem to pose a problem for the thesis that the association of afferent stimuli constitutes the data points. How can we perform the searching and question answering behaviour essential to this kind of feedback control (resonance) without requiring the use of anything but the association of afferent stimuli?

The answer is deceivingly simple: understanding, intelligibility and higher order imagery is required only because without those, we could not store all the required data. Though the higher order relations are themselves intelligible, the lower order associations which they represent are not. Dunker maintains that every intelligible relation is reducible to some general law which is itself unintelligible (1935 p.5, 48, 58-65). Thus we may recognize the fact that blowing on a dying fire will rekindle it, but that fact itself is unintelligible: it simply consists in the association of observables. Even if we apply chemistry to make the association intelligible, and even if we reduce the principles of chemistry to atomic physics, the principles of the latter must be accepted as mere fact.

Dunker further maintains that problem solving depends ultimately on these unintelligible relations, and that the understanding, or the intelligible, often serve only pragmatic purposes. Understanding will guide our choice of reformulations, and therefore restrict an otherwise random trial and error process. The chimpanzee 'knows' that it must search for a stick of appropriate length, not appropriate colour. But beyond the pragmatic value of the intelligible, our solution processes, and thinking in general, relies in the end only on abstracting induction.

> This much is beyond doubt: even in a world of totally unintelligible connections, a subject can learn what matters in each case and can solve new problems on the basis of such general experiences (p.66)

This idea is very Humean (1748). Only the experience of constant conjunction can afford the familiar understanding of necessary connection. That is, the intelligibility of causal relations depends on the repeated association of stimuli.

It appears then, that this idea of necessary connection

> among events arises from a number of similar instances which occur of the constant conjunction of these events. (Section VII Part II)

Now the point that the association of stimuli is what our reasonings, judgements, and decisions are based on can be arrived at from another angle. Dunker argues that problem solving generally occurs through a search of the general connection structures. These connection structures are learned through abstracting induction, and store 'ground-consequence' relations. Now Dunker is somewhat ambiguous on the question of just what form a ground-consequence relation takes. But he suggests that often the relation is a kind of 'leading to...'. That is, events a,b,c,... will lead to another event x. Similarly, Kelley (1971) characterizes causal schemata as "an assumed pattern of data in a complete analysis of variance framework" (p.152). For Kelley, the data stored in a schema pertains to the probability of an effect given a combination of antecedent causes. These schemata are then applied in judgement and reasoning. Here, the information structure has been construed much more broadly; the entire schematic network is involved in reasoning. There is an obvious explanation for this discrepancy. My construal of the schematic network as possibly including any imaginable relation between any set of images is a function of the fact that I take the construal from introspection. These researchers, by contrast, take their construal from external data derived from experiments. As such, they are positing the underlying knowledge structures based on reasoning behaviour. It is therefore fully appropriate to say that their structures are sufficient for such behaviour (assuming that their research is valid). Thus whatever content of the entire schematic network which does not participate in Dunker's connection structures or Kelley's causal schemata can be seen as unnecessary for the kinds of reasoning examined by these authors. Furthermore, the content of the connection structures or causal schemata is nothing more than what can be learned by induction. That is, the association of afferent stimuli. (The general form of such structures will in the next section be called "predictive content".)

In this section, an effort was made to establish the point that the association of afferent stimuli is sufficient data on which to base our behaviour and reasoning. The main idea is that these associations are what is required for the adequate functioning of the organism. Beyond that, higher order relations serve mainly the efficiency of storage and recall. In the next section, these conclusions are applied to scientific investigation and theory. The analog of the association of stimuli in science is the correlation of observables. It will be argued that these serve as the data points, and theory which makes claims beyond the data is confabulated.

Section 7: Psychological Empiricism

The thesis of psychological empiricism is that the predictive content of our semantic theories (including scientific theories) lies in the association of stimuli (relation of observables); this constitutes the data points between which gaps are filled by confabulation. This thesis provides an approximate line between what is data and what is confabulation, and thus a way of rating the epistemic grounding of our various knowledge claims.

This section is devoted to an explanation of psychological empiricism. I will answer some common objections offered against similar theses, and include some reasons to accept this one. I do not intend to prove the truth of my views, but only their consistency and plausibility. That is, my arguments are illustrative, not demonstrative. We should be clear from the start that the question is begged; I assume the thesis of psychological empiricism and show how the rest of the pieces should fall to gain consistency. This is important since my argument involves the use of some unobservables (goal states, causal mechanisms), and since it demonstrates that unobservables must be confabulated.

In the last section, my goal was exactly that. If I am right in that human behaviour is best explained by the mechanisms posited in this paper, and in that these mechanisms do not require any more correspondence with reality than is contained in the association of stimuli, then human behaviour can be explained on the thesis of psychological empiricism.

Now science too is a part of human behaviour, and thus we have reason to be sceptical of the unobservables posited there, classifying them as (psychologically and pragmatically useful) confabulation. However, that argument seems too quick. Science seems too far detached from the kinds of behaviour on which my arguments were based (driving or memorizing lists of words or abstract problem solving or deciding what to eat). Such considerations are what motivates this section. Furthermore, the philosophy of science provides an appropriate and familiar context for explication.

Two things are required before psychological empiricism can get off the ground. First is an observation-theory dichotomy, and second is a clear notion of predictive content.

The observation-theory dichotomy

What we want here is a clear way of including dogs, trucks and time as observable, while excluding electrons, causes and gods. There are various ways in which this distinction will not work. For instance, we cannot claim that the former are directly perceivable while the latter must be inferred. This is because perception is always mediated by our conceptual schematic network. By the time a pattern of retinal excitation or tympanic vibration becomes a fully formed image, it is no longer anything basic. This process is a case of translation, and that is dependant on classification and categorization, *i.e.*, inference.

As retinal stimulation is processed through the optic nerve, lateral geniculate nucleus, Broadman's area 17, 18 and 19, *etc.*, the original information becomes increasingly theoretic. It may finally end in a fully conceptual image of an angry god who requires the sacrifice of two goats, three chickens, and an ox. The problem is where to draw the line.

We can put the question as follows: what properties and objects are transduced by the visual system, to which properties and objects does the visual system respond? The answer will depend on what we include in our description of the visual system (Fodor and Pylyshyn 1981). If it is the entire organism, then anything imaginable is transducible. If it is only the retina, then not even trucks are included. The pattern of excitation on the retina does not by itself contain any information about trucks; to infer that such an obviously observable object is in the visual field requires higher level theory about trucks and what they look like. If we draw the line anywhere in between, we must do it arbitrarily.

The same point has been by Grover Maxwell (1962). He argues that there is a continuum between cases of so called 'pure' observation and cases of theory-laden observation. In his example, it begins with looking through a vacuum and continues through "looking through a windowpane, looking through a glasses, looking through binoculars, looking through a low-power microscope, looking through a high-power microscope, *etc.*".

It should also be noted that sense data cannot serve our purpose. It is by no means clear that sense data are atheoretic. As stated a few paragraphs back (and in section 2), all the contents of imagery are to some extent conceptual. Even if phenomenological sense data were the basis of our epistemology so that all claims must be justified by that alone, then it seems that knowledge of the external world would be impossible.

So how then can we make the required distinction? Evidently, the concept (observable) is too vague to draw a clean, non-arbitrary line. However, it does not follow that there is no difference between observables and unobservables. Following Van Fraassen (1980), we can describe a spectrum between the two sorts by providing paradigm cases of each. This is fairly intuitive, since it is likely that even the hard core scientific realists would agree that trucks are more observable than electrons, and electrons are more observable than gods and demons.

Anyone who would agree that the impossibility of verifying the existence of gods and demons provides some grounds for denying their reality, should see my motivation for denying the reality of causation, electrons, and unobservables generally. (I will have more to say on this point in my discussion of abduction.) Thus viewed, it seems that much of the dispute between the scientific realist and the psychological empiricist rests in where on the spectrum they draw the line. In this context then, my major claim is that the line should be shifted a little to the left, toward the observable. If I have shown nothing else in the preceding pages, I have at least shown that confabulation is far more prevalent than commonly thought, and that it can serve important psychological and pragmatic functions. This alone should provide some reason for shifting the line.

There is another point of Maxwell's (1962) which is important in this context. He argues that what is classified as observable is dependent on the physiology of humans, and thus accidental. Surely this is true, were we much smaller, then atoms might have been included in our observational vocabulary. His point is that such accidental considerations can have no bearing on ontology. Whatever the fundamental building blocks of the universe are is not dependant on the structure of our sensory apparatus. In this I completely agree. But what then of my scepticism toward the existence of electrons? This was only an epistemological point: claims about unobservables are less well epistemically grounded than claims about observables, and thus more likely to be confabulation.

As it stands, the dichotomy lacks a definite analysis. This is intentional; it avoids some of the complexities and counterexamples which would arise given a precise distinction. For example, we might say that

> X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it. (Van Fraassen 1980 p.16)

But this opens up the possibility that the creation of the earth by God be classified as observable since were I here six thousand years ago, I would have witnessed it. While at the same time, we do not want to exclude every event for which the circumstances for its observation have passed. Consider C.B. Martin's (1984) example concerning a salt crystal which has been dissolved in a saline solution. It is now too late to observe the weight of that crystal, yet we still want to classify that weight as something observable. What we want to rule out is those events and entities which must be inferred from the observables, those which are not themselves observable.

Again, since even the simplest objects of imagery must be inferred from something more basic, we cannot draw a clear line. And again, neither the retinal stimulations (or its analogous counterpart in the other sense modalities) nor the phenomenological sense data (which may constitute the 'something more basic') will suffice for our purposes. Thus we are left with intuitive judgements of comparative observability to construct the spectrum between observable and theoretical. Though perhaps not ideal, it will prove sufficient.

The tripartition of theory

The content of any semantic theory (scientific theories in particular) can be divided into three parts: the predictive content, the empirical content, and the remainder. The predictive content is (generally properly) contained in the empirical content, which in turn is (generally properly) contained in the entire content. I will employ a standard underdetermination argument to show that whatever content is not empirical is confabulated, then argue that the conclusions of section 6 suggest that there is good reason to classify whatever empirical content is not predictive as confabulation too. The latter claim forms the heart of psychological empiricism.

The distinctions are relatively straightforward. The empirical content is characterized by a list of all the observable consequences of a theory. This is close to co-extensive with the verificationists' criteria of meaning. Here, we avoid three of the more striking objections against that criteria. First, the criteria often refers to those objects or events which are 'in principle verifiable' or 'observable in principle'. As such, they might leave out such things as the weight of the salt crystal in C.B. Martin's example just discussed. Thus, by leaving the analysis of 'observable' vague and intuitive, we can capture the essence of the distinction without dealing with purported counterexamples.

Second, it has been charged that the empirical content of a theory does not truly capture the meaning of that theory. It is argued that two theories with identical empirical content do not really mean the same thing when one posits entities which the other does not. Compare the theory that God is causally responsible for the observed regularities in nature, with the theory that it is the behaviour of unobservable particles which serve as explanation for the observed phenomena. It is *prima facie* obvious that these two theories do have different meanings: the first posits God, the second does not. I am fully willing to embrace this conclusion. The meaning of a theory is captured only by its entire content. I suggested in section 3 that the meaning of an image should best be understood as its semantic situation, *i.e.*, as all its various relations to other images. Since a semantic theory is just a collection of such related images, its meaning too should be understood as comprised by its internal and external relations. Since images can be of unobservables (of anything imaginable really) those too contribute to the meaning of theories. These considerations have no bearing on the plausibility of distinguishing empirical content from the rest.

Third, it is often the case that a theory does not deductively imply its empirical content. This makes the notion of "observable consequences" (as used in my definition) suspect. If it is required that the empirical content be derivable from the entire content, then some theories will have no empirical content at all. I therefore suggest the following revision. The empirical content of a theory is characterized by a complete list of the likelihoods it ascribes to each possible observation, *i.e.*, the probability of each observable outcome in specific circumstances (also in terms of observables), given that the theory is true. The form would be as follows: when we observe that circumstances of type a,b,c,... obtain at a time t, then we expect an observable outcome of type x to obtain at time t' with probability Y. This account will allow us to equate the empirical equivalence of theories with their ascription of identical likelihoods to every observation, and thus to run a very clean analysis of the classical underdetermination argument. That will have to wait for the next subsection though.

The predictive content of a theory contains only what has actually been observed. Its sentences take the form: when observable circumstances of type a,b,c,... obtain at a time t, then outcome x obtains at time t' in w out of z trials. Now, it is the thesis of psychological empiricism that sentences of the empirical content, when justified, follow inductively from sentences of the predictive content (generally with Y = w/z).

Consider a simple example. A particularly naive person observes that sea water in an aluminum pot will boil if placed on a fire. Now she induces that tapwater in a stainless steel pot will boil if placed on a stove element which has been turned on. To make this prediction, she must have guessed that both kinds of water, pots, and heat sources have the relevant properties in common, *i.e.*, that the items in question were members of the same natural kind.

My point is that, though the example was oversimplified, this kind of inductive inference is the basis of all of our knowledge. Whatever empirical content a theory has which cannot be induced from the predictive content is confabulated. We should expect then, that predictions not thus justified will generally not be confirmed. I will discuss the objection that such predictions are sometimes verified in the subsection on surprising predictions. For now, let us look at the positive argument.

I argued in the previous section that the only information required for the successful application of the various psychological mechanisms is the association of stimuli. The analog of those associations in the case of scientific theories is the predictive content. In the psychological case, we learn to associate afferent signals of various kinds. To apply these in (relatively) novel situations, and to learn them in the first place, we must rely on the uniformity of nature. That is, were there no such uniformity, the re-afference specific to a perceptual situation would not consistently be rewarding, both in the case of learning and in the case of later application of what is learned. The point is that we must use induction on the associations of stimuli both to learn them and to depend on them once learned. This is only to establish the analogy between the psychological case and the scientific case: in both, we induce on the regularities of stimuli/ observables to make the predictions required for behaviour and technology. Just as induction on the associations of stimuli is sufficient for behaviour in the psychological case, induction on the relations of observables (predictive content) is sufficient for prediction and technology development in the scientific case.

Recall that I have made no claim to a knockdown argument for this thesis. I know of no way to prove either that the association of stimuli constitutes all of our knowledge in the psychological case, or that the predictive content as characterized constitutes all the information required for the successes of science. Recall that I have assumed this, and attempted to illustrate its plausibility and consistency, as well as provide an answer to some purported objections (in the next few subsections). Thus viewed, the analogy in the preceding paragraph serves only to provide a motivation for psychological empiricism, not to establish it as irrefutable fact. The main idea was just that since science is an aspect of human behaviour, if behaviour can be adequately served by the association of stimuli, then science can be conducted successfully by induction on the predictive content.

Now this conclusion is not quite right. The fact is that science would probably not go very far if scientists learned only what is contained in the predictive content. Huge lists of observed phenomena could not be efficiently memorized and understood. This is the pragmatic role of theory. Just as confabulation served to aid memory in the psychological case (as well as understanding and problem solving indirectly), it serves similar purposes in science. Only by giving a student a comprehensible theory can we motivate him to explore various consequences and predictions. Thus we can see how confabulation is pragmatically useful in science too. Theoretical entities provide a framework for memorizing the observed regularities of our measuring instruments. Simplicity of theory provides for efficiency of storage. Coherence and generality allow individual theories to be incorporated into one large schematic network. All these are pragmatic virtues of theories. Without them, and without theory and confabulation generally, scientific progress would be considerably slower than it actually is. But notice that the pragmatic usefulness of such features is dependent on the physical makeup of human beings. Were our memory and reasoning capacities unlimited and unerring, we could dispense with everything but the predictive content. As historical accidents, these features of scientific theories have no bearing at all on the theory's truth. Coherence with previously held theory, for example, is a good criterion for theory acceptance because of its psychological advantages, but not because a theory which coheres is more likely to correspond. This is particularly vivid when the theories in question are empirically equivalent.

It should be clear from the preceding paragraph that I do not intend the foregoing to be a prescription for scientists. A reduction of theory into its predictive content would be detrimental to scientific progress. The conclusions are for epistemology only: we should be wary of theories insofar as their content goes beyond the empirical content, and we should be wary of that part of the empirical content not justified by induction on the predictive content.

Now it should be noted that not all predictions induced from what has been observed will be accurate. Had our naive water boiler been equipped with a thermometer, she might have predicted that the tapwater would boil at the same temperature as the seawater; she would be unlikely to have guessed that covering the pots or changing the altitude effects the outcome. Knowledge of the relevant factors must be acquired empirically. I claim that predictions which can be justified by induction on what has been observed are better epistemically grounded, not that they are infallible.

We should now be able to see why the claim that "My truck goes from zero to sixty in seven seconds" is better justified than the claim that "the electron spins with the value -1/2". the former claim can be justified by induction on predictive content. The experimental setup is described in terms of the truck, the speedometer, the clock, the slope and friction of the road, the wind velocity, *etc.* These are all observable, and the claim does not go beyond induction on the experimental results. By contrast, the claim about the electron cannot be induced from predictive content, since neither electrons nor electron spin can enter into sentences of that content (since they are not observable). Rather, "spin" must be cashed out in terms of the measurement of the magnetic field produced, and "electron" must be cashed out in terms of cloud chambers and the like. To go from the observation of correlations of such complex measurement devices to the idea of a particle which spins like a baseball is indeed a large leap. I suggest that epistemology should recognize such leaps as less well justified than observed statements about trucks.

Underdetermination and miracles

The standard anti-realist underdetermination argument runs something like the following (based on Boyd 1983). Let T be some theory which postulates unobservable entities and/

or causes. We can always construct another theory T', with identical empirical content but with a contradictory account of the unobservables entities and/ or causes. Since scientific confirmation or disconfirmation always consists in the verification or refutation of a theory's observable consequences, it can never afford any reason to prefer T over T', or *vice versa*.

Accordingly, knowledge of unobservables is impossible since scientific (observable) evidence does not bear on the question of which account of the unobservables is correct. Further, T' might simply take the Vailhingeresque form: all of the observables are exactly 'as if' the unobservables of T existed and behaved exactly as prescribed, except that they don't exist. As such, the truth of the unobservable part of T (the entire content minus the empirical content) is inevitably highly questionable.

The problem is, if you buy this argument form, then it seems you are committed to wholesale scepticism. Underdetermination arguments proceed from what is not directly empirically testable to what is unknowable or 'highly questionable'. Consider Putnam's brains in a vat, Descartes' evil demon hypothesis, or Russell's proposal that the universe was created five minutes ago. The important feature of all these examples is that there is no empirical evidence which could refute them.

What is required here is a way to make the underdetermination argument run so that the unobservable entities and causes of science are discredited, but the immunity of induction on observables is maintained. There does not, however, seem to be any special feature of the latter which can afford it such immunity, outside of completely subjective and arbitrary considerations. There is no objective justification for selective application of the underdetermination argument. But perhaps that is not really required after all. Recall that my claim was that the predictive content of theories and the predictions induced from it are better epistemically justified than the non-empirical content and the empirical content which that supposedly justifies, not that the former provides any kind of certainty. This claim can be borne out while maintaining the conclusions of the underdetermination argument.

Recall that two theories are empirically equivalent exactly when they have the same empirical content. That is, when they ascribe identical likelihoods to every possible observation. Formally, T and T' are empirically equivalent if and only if for every possible observation O, $P(O/T) = P(O/T')$. The underdetermination argument relies on this to show that empirical evidence alone is insufficient to distinguish between empirically equivalent theories.

However, if we adopt the canons of Bayesian reasoning then there is something we can turn to resolve questions between empirically equivalent theories: prior degrees of belief. Now these are simply subjective attributions of probability to various hypotheses, and as such won't offer any objective justification for preferring T over T' or *vice versa*. Taken at face value, they simply dictate that whichever seems *a priori* more plausible to us is more likely to be true. That kind of reasoning is highly suspect. Why should our subjective plausibility ratings be trusted as a guide to truth? Even if they could, this would not help resolve the issue in question. Though everyone might agree that it is more plausible that we are people in a world than brains in a vat, there will be disagreements over the status of the unobservables of science. However, there is a way to take advantage of this inevitable subjectivity to cast doubt on the unobservables while maintaining the good (better) standing of induction on observables.

Let us begin by dividing cases of underdetermination into three types. First is what we'll call "wholesale underdetermination". These are cases like the brains in a vat and the evil demon. If we run underdetermination arguments on these cases, wholesale scepticism results. I suggest that we use priors to construct an abductive escape. Since my being a person in the world is a better (*a priori* more plausible) explanation for my phenomenal experiences than my being a brain in a vat, I conclude that indeed I am the former. There is an added bonus to this abduction. We can build in to the notions of "person" and "world" the idea that the world is inherently uniform, and that people take advantage of the resulting regularities for prediction and behaviour. Thus we build in the idea that induction works when applied properly.

The next kind of underdetermination is called "Goodman underdetermination". It is just a restatement of Nelson Goodman's (1955) now classic problem of projectibility. Roughly, the problem proceeds by showing that for every intuitively plausible inductive inference, there are infinitely many unintuitive and empirically inconsistent rivals for which the evidence may be equally confirmatory. For example, suppose every emerald examined for colour has been found to be green. Then we are well advised, provided certain conditions are met, to project the hypothesis that all emeralds are green, and well justified in predicting that the next emerald we examine for colour will be green. But, the story goes, every emerald hitherto examined has also been grue (something is grue if it has been examined before time t and is green, or if it has not been examined before t and is blue). Thus, now being t, aren't we well advised to project the hypothesis that all emeralds are grue, and to therefore predict that the next emerald we examine for colour will be blue? If we accept only past evidence into our considerations, then the two predictions must have the same probability. (The problem is not completely analogous to the underdetermination argument as set out here. The competing hypotheses in this case are not empirically equivalent, they must always conflict on at least one observation if the problem is to be applied. It is still an example of underdetermination though, what is underdetermined is not the unobservables, but the predictions. This is, in effect, the classic curve-fitting problem.)

The canons of confirmation theory do not afford either a clear way to distinguish lawlike hypotheses (legitimate candidates for induction) from accidental generalizations, or a justification for the classification. Goodman proposes and rejects at least six possibilities before finally accepting a subjective solution. In effect, Goodman suggests that we turn to the past predictive success of our predicates and resolve conflicts between hypotheses by projecting the one whose predicates have been used with encouraging results more frequently. That is, we trust the predictions of the green hypothesis over the grue, because of its "more impressive biography" (p.94). This amounts to saying "just keep doing what you're doing, it's working".

Goodman's solution relies on our language use and our accidental history, this is why I classify it as subjective. The Bayesian solution (as exemplified by Sober 1994) is similarly subjective. He suggests we must rely on our prior degrees of belief in the plausibility of each hypothesis to distinguish them.

"Underdetermination of unobservables" is the third type of case. This type was already discussed at the start of this subsection. Again, the underdetermination argument provides good reason to be sceptical of unobservables, and again subjective considerations can provide a way out. Putnam (1975), for example, has provided an abductive argument for realism about unobservables.

> The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle. That terms in mature scientific theories typically refer... that the theories accepted in a mature science are typically approximately true, that the same term can refer to the same thing even when it occurs in different theories- these statements are viewed by the scientific realist not as necessary truths but as part of the only scientific explanation of the success of science. (p.73)

Before discussing this argument directly, let us look at how it is possible for the empiricist to accept subjective considerations in the first two cases, but outlaw them in the third. There are obvious parallels between the three cases.

The first thing to notice is that this is exactly what the empiricist needs to do. We cannot say that induction on observables is objectively justified, yet if we allow subjectivity to enter there, how can we run underdetermination arguments against unobservables? We need two kinds of abductive arguments to justify induction on observables, and with each comes some degree of subjectivity and hence some degree of scepticism. Hume (1748) taught us that induction can be neither deductively or inductively justified without circularity. That in conjunction with wholesale underdetermination is sufficient to demonstrate that subjectivity is inherent in the use of inductive inference. That is, without abduction, it seems impossible to demonstrate that we are in a world in which induction works. Next, Goodman underdetermination shows that subjectivity is required again for each individual inductive inference. There are no perfectly objective standards for the selection of projectible predicates or lawlike hypotheses. Thus subjectivity enters twice into the empiricist's justification of induction on observables.

The realist, by contrast, must allow both these kinds of subjective inference, as well as a third abductive step. The latter is evident in Putnam's miracle argument. So while empiricist justifications require the rejection of the first two kinds of underdetermination by subjective means, the realist requires the rejection of the third kind as well. If we allow that each time subjectivity enters into our epistemic justifications a degree of scepticism about those justifications inevitably tags along, and if we agree that the scepticism is compounded with the reliance on subjectivity at each step, then it is easy to see why predictions induced from observation are better epistemically justified than predictions justified by unobservable theoretical entities. The empiricist justifications admit of a certain degree of scepticism. The realist justifications must admit scepticism to this same degree (since they too accept induction on observation) plus whatever scepticism results from the abduction to unobservables. Thus the scepticism inherent in empiricism is always less than the scepticism inherent in the realist picture.

This argument was motivated by Boyd's analysis of Fine's objection to the miracle argument. Fine (1984) pointed out that the miracle argument is abductive in form, and hence not convincing since abduction is exactly what is at issue between the realist and the anti-realist. Fine's anti-realist accepts induction on observables as a guide to truth, but rejects abduction and therefore theoretical entities and causes. (The underdetermination argument establishes that abduction is probably the only way to infer the existence of unobservables.) Thus the miracle argument is seen as employing an argument form which the anti-realist does not accept. This is similar to demonstrating the falsity of mathematical intuitionism by *reductio ad absurdum*. Though valid to the Platonist, the argument will never convince an intuitionist since it employs an argument form they explicitly reject. In response to this, Boyd (1991) has very correctly pointed out that Fine's anti-realist is not a wholesale sceptic and so is committed to abduction in justifying induction on observables.

The argument presented here is a response to Boyd. It is true that induction on observables is not immune to scepticism, and that the anti-realist must accept at least some abductive arguments. But it is also true that the anti-realist needs accept fewer such arguments than the realist. Since with each abduction comes subjectivity and scepticism, the anti-realist's justifications are better epistemically grounded (less subject to scepticism) than the realist's. This has been my claim all along.

There is more to the essence of the miracle argument than its form. It presents a challenge to the empiricist. How can we make sense of the successes of science if not through realism? We may take the "successes of science" to mean two things. First is the technology science affords, and second is the success of predictions based on evidence from a seemingly different domain of inquiry. Putnam wants to claim that realism is the only coherent way to explain these. I will conclude this subsection with an alternative explanation of the former, and address the later in the next.

The central claim of psychological empiricism was that the predictive content of theories is contained in observation, and that non-empirical content serves psychological functions only. In this context, the function of unobservable entities is to collect the predictive content into an efficiently stored and recalled schemata. The arguments of the last section were purported to illustrate how psychological mechanisms could function adequately if data were construed as afferent stimuli. Recall that my claim was not demonstrated deductively, but that I admitted that the question was begged, and endeavoured to explain human behaviour on the hypothesis.

Now, in effect, what I have done is reject the realist's call for explanation. That is, I reject the use of subjectivity or abduction to avoid the sceptic's claim of the underdetermination of observables because I do not agree that unobservables are real. What does not need explanation, then, is our knowledge of unobservables (we don't have any). However, the essence of the miracle argument is not captured in this move. What Putnam claims is that the success of science needs explaining. This motivation for realism will stand despite any philosophical arguments against unobservables. Thus my argument might be construed, not as rejecting the need for explanation, but as offering an alternative explanation. Where Putnam advances scientific realism as the best explanation for the successes of science, I advance psychological empiricism as an alternative explanation, and claim that it is better. The psychological system which requires only the association of observables as data constitutes the alternative explanation. That is, if all the data required to reason and function adequately is contained in the association of observables, then unobservables need not refer to anything for science to function adequately. This alternative should remove the motivation for realism expressed by the miracle argument.

Surprising predictions

Now predictions can be divided, roughly, into two kinds. The first is that kind which can be justified by induction on observation. The second kind cannot be similarly justified. This second kind we'll call "surprising". Such predictions generally arise from the non-empirical content of theories. They may be cases where the prediction is 'justified' by considerations about unobservable causes and entities, or cases where a theory is applied too far from the domain where it was conceived. In either case, the important thing about surprising predictions is that they are not justified by induction on the predictive content of the theory.

When scientists theorize, they often go further than simply collecting the observations. Newtonian mechanics is an example of a theory which does not. Atomic theory is one which goes far beyond what is justifiable by its predictive content. The verification of surprising predictions then poses a problem for the anti-realist. Psychological empiricism claims that successful predictions are so because they are either explicitly justified or implicitly justifiable by induction on observables. Thus when predictions 'justified' by higher order theoretical considerations are verified, and when empirical content not inducible from predictive content is confirmed, the psychological empiricist cannot explain that success. It seems that realism about the non-empirical content offers the only explanation. I intend to argue that the success of such predictions is explicable by chance.

What we want is a way to assess the probability that a surprising prediction should come out true. There are a number of factors which must enter into consideration. First, any scientific prediction must be observable. Though sometimes couched in terms of unobservables, a quick translation of the methods section of any journal article will yield a description of the observables for that case. These may include things like electron guns and particle accelerators (considered as wholes: described the way you would describe it to a mechanic or engineer who was going to built one for you, not the way you would describe it to a student), as well as measurement devices like geiger counters and photoelectric plates. The point is that the prediction can always be given in terms of the observable part of the experimental setup as well as an interval on a measurement instrument which will constitute confirmation. Suppose this instrument has a pointer, and we take as confirmation a pointer reading between 15 and 18. Now presumably, that meter will only be able to register quantities in a finite interval (say, between 0 and 50). Thus, even if we were completely naive and predicted haphazardly, there is always some chance that our prediction will be confirmed.

Add to this initial probability the fact that though some surprising predictions have been verified, some also have not. The idea is that the confirmation of special relativity and Rutherford's model of the atom are counterbalanced by phlogiston, the luminous ether and the fact that Leeuwenhoek saw fully formed human beings swimming around in his semen. Now to show that surprising predictions are due solely to chance there must be many more which were unsuccessful than were confirmed. This is necessary because of the presumption that the initial probability established in the previous paragraph was rather small. The ratio of successful to unsuccessful surprising predictions can only be borne out by a thorough search of the history of science. If the reader is not convinced that such a count will yield an outcome sufficiently in my favour, consider the following.

The final factor to enter the consideration is human biases. I will paint a picture of the scientist as a hypothesis tester who searches almost exclusively for confirming evidence, who accepts confirming evidence casually while critically evaluating disconfirming evidence, whose theory can survive the total discrediting of its evidence base, and who places more weight on early evidence. This, in conjunction with the initial probability and the ratio of confirmed surprising predictions might substantiate the claim chance is indeed the culprit.

Throughout this paper an image of the organism as a hypothesis tester has been stressed. The difference between testing hypotheses and gathering evidence is the difference between top-down and bottom-up processes. Hypothesis testing is expectation driven. We are not simple processing information, we are always 'searching for something'. This tendency can be found everywhere in the human: it is the same gap-filling process which yields confabulation, it is the model of search (the reference signal) which we employ in problem solving and conditioning, it is the active keyhole. The human as hypothesis tester is popular in the psychological literature, and the scientist as hypothesis tester is also seen in the philosophical literature (Kuhn 1962, Popper 1963).

The confirmation bias was thoroughly researched by Wason. For example, in a 1966 study subjects were asked to determine whether the following rule is true or false: "If a card has a vowel on one side, then it has an even number on the other side". Four cards were displayed face up reading E, K, 4, 7. Only 10% chose the card which would offer what may be construed as disconfirming evidence (the 7 card), while every subject chose to search for the confirming evidence (the E card). In 1960, he had subjects try to determine the rule governing a sequence of natural numbers. They provided the subjects with an initial example which conformed to the rule (2 4 6) then asked them to generate possibilities which would then be classified by the experimenter as positive (fitting the rule) or negative. Before each sequence offered, the subject was to write down the hypothesis being tested. He found a marked tendency to ignore the falsification strategy, even though it is the best for such circumstances. Some of the evidence for the confirmation bias involves the analysis of covariation and is not applicable to scientists who employ statistical tools. Scientists do, however, determine which experiments are run, and thus a confirmation bias can yield some added plausibility to my claim that confirmed surprising predictions are explained by chance.

In a brilliant study by Lord, Ross, and Lepper (1979), subjects were chosen based on their reported attitude toward capital punishment. One group was pro, the other was con. The subjects were given brief reports of studies which either supported or rejected the deterrent effects of capital punishment, then the experimenter discussed the strengths and weaknesses of each study in detail. The results were unambiguous. Subjects in each group reported the study favouring their position as better conducted and more convincing. Some in each group offered the discussed strengths of the study they favoured when asked for written comments, while offering criticism of the other study. Further, the overall result of the studies and discussion was polarization. Each group left the experiment more secure in their initial beliefs, though they were exposed to the same evidence.

Belief perseverance is the retaining of a belief in spite of new evidence which disconfirms it, or in spite of the discrediting of the evidence on which the belief was initially formed. Francis Bacon recognized this phenomenon (1620 XLVI).

> The human understanding when it has once adopted an opinion... draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects, in order that by this great and pernicious predetermination the authority of its former conclusions may remain inviolate.

Ross, Lepper, and Hubbard (1975) showed that beliefs can survive the total discrediting of their evidence base. Subjects solved problems and were given immediate feedback (right or wrong). The number of successes was determined by the experimenters in advance, thus the subjects' actual performance had no bearing on their feedback. After the debriefing, where the subjects were told of the deception and given the paper which determined their feedback before they arrived, only three out of twenty subjects in the success condition (24 out of 25 correct) though their scores were worse than average, while only three out of twenty failure subjects (10 out of 25 correct) thought that their actual score was better than average. Similar experiments have not been run due to ethical concerns about leaving lasting impressions on subjects. However, in the 1960 Wason study, it was found that more than 50% of sequences offered by subjects immediately following the disconfirmation of their hypothesis was still consistent with that hypothesis. This phenomena of belief perseverance is widespread. Wason and Johnson-Laird (1972) offer this and other evidence in support of Kuhn (1962) who arrived, quite independently, at an observation of the same tendency in scientists.

> Though they may begin to lose faith and then to consider alternatives, they do not renounce the paradigm that has led them into crisis. They do not, that is, treat anomalies as counterinstances, though in the vocabulary of philosophy of science that is what they are. (p.77)

Belief perseverance can be seen as a function of hypothesis testing. Scientists will adhere to their paradigm despite the fact that such adherence is no longer justifiable by the evidence. They will adhere until a new paradigm is in place. They do not research in a vacuum, they are always testing hypotheses.

The tendency to place more weight on early evidence is known as the primacy effect. In a thorough review of order effects, Jones and Goethals (1971) attributed some cases of primacy to the hypotheses being tested and the expectancies produced thereby. In this assimilation of new evidence, information is distorted to fit the mould of preconceptions, provided the information is not too discrepant (contrast effects).

There is therefore a parallel between hypothesis testing and resonance. We are not just 'trying on' a belief (system) to see if it works, the hypothesis itself creates an expectancy. It makes it more probable that the input is within the reference interval. It can create the satiating input out of the data. That is, to an extent, we create our own reality.

Works Cited

Ashcraft, Mark H. (1994). Human Memory and Cognition. 2nd Edition. New York: HarperCollins College Publishers.

Bacon, Francis (1620). The New Organon.

Bartlett, Sir Frederic (1932). Remembering. 3rd Edition. Cambridge: Cambridge University Press, 1954.

Bechara A., H. Damasio, D. Tranel and A. Damasio (1997). "Deciding Advantageously Before Knowing the Advantageous Strategy." *Science* 275: 1293-1295.

Bem, Daryl (1972). "Self-Perception Theory." In L. Berkowitz (Ed.) Advances in Experimental Social Psychology Vol 6. New York: Academic Press.

Berlyne, N (1972). "Confabulation." *British Journal of Psychiatry* 120: 31-39.

Bousfield, W.A. (1953). "The Occurrence of Clustering in the Recall of Randomly Arranged Associates." *Journal of General Psychology* 49: 229-240.

Bower, G.H. (1970). "Analysis of a Mnemonic Device." *American Scientist* 58, 496-510.

Boyd, Richard (1991). "On the Current Status of Scientific Realism". in Boyd *et al.* (Eds.) The Philosophy of Science. Cambridge: MIT Press.

Casscells, W. et al. (1978). "Interpretation by Physicians of Clinical Laboratory Results." *New England Journal of Medicine* 299.

Conrad, R. (1964). "Acoustic Confusion in Immediate Memory." *Journal of Experimental Psychology* 55: 75-84.

Cosmides, Leda and John Tooby (1996). "Are Humans Good Intuitive Statisticians After All?" *Cognition* 58: 1-73.

Dunker, Karl (1935). On Problem Solving. Trans. Lynne S.

Lees. *Psychological Monographs* Vol.58, No.5: 1945.

Edwards, Derek and Johnathan Potter (1992). Discursive Psychology. London: Sage.

Einhorn, H.J. and Robin Hogarth (1978). "Confidence in Judgement: Persistence of the Illusion of Validity". *Psychological Review* Vol.85, No.5: 395-416.

Ferster, C.B. and B.F. Skinner (1957). Schedules of Reinforcement. New York: Appleton-Century-Crofts.

Festinger, Leon and J.M. Carlsmith (1959). "Cognitive Consequences of Forced Compliance." *Journal of Abnormal and Social Psychology* 58: 203-210.

Fine, A. (1984). "The Natural Ontological Attitude". in J. Leplin (Ed.) Scientific Realism. Berkeley: University of California Press.

Fischer, Richard et al. (1995). "Neuropsychological and Neuroanatomical Correlates of Confabulation." *Journal of Clinical and Experimental Neuropsychology* Vol.17, No.1: 20- 28.

Fodor, J.A. and Z.W. Pylyshyn (1981). "How Direct is Visual Perception". *Cognition* 9: 139-196.

Gazzaniga, M.S. and J.E. Ledoux (1978) The Integrated Mind. New York: Plenum Press.

Goodman, Nelson (1953). Fact Fiction and Forecast. 4th Edition. Cambridge: Harvard, 1983.

Hershberger, W.A. (1990). "Control Theory and Learning Theory." *American Behavioral Scientist* Vol.34, No.1: 55-66.

Hume, David (1748). An Enquiry Concerning Human Understanding.

Kuhn, Thomas S. (1962). The Structure of Scientific Revolutions. Chicago: University of Chicago Press

James, William (1890). Principles of Psychology. New York: Holt.

Jones E.E., and G.R. Goethals (1971). "Order Effects in Impression Formation". in E.E. Jones *et al.* (Eds.) Attribution: Perceiving the Causes of Behaviour. Morristown: General Learning Press, 27-46.

Joseph, Rhawn (1982). "The Neuropsychology of Development: Hemispheric Laterality, Limbic Language, and the Origin of Thought." *Journal of Clinical Psychology* Vol.38, No.1: 4-33

Joseph, Rhawn (1986). "Confabulation and Delusional Denial: Frontal Lobe and Lateralized Influences." *Journal of Clinical Psychology* Vol.42, No.3: 507-520.

Joseph, Rhawn (1996). Neuropsychiatry, Neuropsychology, and Clinical Neuroscience. 2nd Edition. Baltimore: Williams and Wilkins.

Katona, George (1940). Organizing and Remembering. New York: Columbia University Press.

Kelley, Harold (1971). "Attribution in Social Interaction". in E.E. Jones *et al.* (Eds.) Attribution: Perceiving the Causes of Behaviour. Morristown: General Learning Press, 1-26.

Kopelman, M.D. (1987). "Two Types of Confabulation." *Journal of Neurology, Neurosurgery, and Psychiatry* 50: 1482-1487.

Martin, C.B. (1984). "Anti-Realism and the World's Undoing". *Pacific Philosophical Quarterly* 65: 3-20.

Maxwell, Grover (1962). "The Ontological Status of Theoretical Entities". in Minnesota Studies in the Philosophy of Science. Vol. III. Minneapolis: University of Minnesota Press.

Mayr, Otto (1969). The Origins of Feedback Control. Cambridge: MIT Press.

McDowell, John (1994). Mind and World. Cambridge: Harvard University Press.

Metcalfe, Janet (1986). "Feeling of Knowing in Memory and Problem Solving." *Journal of Experimental Psychology: Learning, Memory, and Cognition* Vol.12, No.2: 288-294.

Mills, C.W. (1940). "Situated Actions and Vocabularies of Motive." *American Sociological Review* Vol. 5: 904-913.

Leahey, T.H. (1991). A History of Modern Psychology. 2nd Ed. New Jersey: Prentice Hall.

Lord, C.G., Lee Ross and Mark Lepper (1979). "Biased Assimilation and Attitude Polarization". *Journal of Personality and Social Psychology*: 2098-2109.

Neisser, Ulrich (1967). Cognitive Psychology. New York: Appleton-Century-Crofts.

Newell, Allen, J.C. Shaw and Herbert A. Simon (1963). "The Process of Creative Thinking." In H.E. Gruber *et. al.* (Eds) Contemporary Approaches to Creative Thinking. New York: Atherton.

Nisbett, Richard and Lee Ross (1980). Human Inference: Strategies and Shortcomings of Social Judgement. New Jersey: Prentice-Hall.

Nisbett, Richard and Timothy Wilson (1977). "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* Vol.84, No.3: 231-259.

Popper, Karl (1963). Conjectures and Refutations. London: Routledge and Kegan Paul.

Powers, W.T. (1973). "Quantitative Analysis of Purposive Systems". *Psychological Review* Vol.85, No.5, 417-435.

Powers, W.T. (1973a). Behavior: The Control of Perception.

Chicago: Aldine.

Powers, W.T. (1973b). "Feedback: Beyond Behaviorism." *Science* 179, 351-356.

Putnam, H. (1978). Meaning and the Moral Sciences. London: Routledge and Kegan Paul.

Ross, Lee and Richard Nisbett (1991). The Person and the Situation. New York: McGraw-Hill.

Ross, Lee, Mark Lepper and Michael Hubbard (1975). "Perseverance in Self-Perception and Social Perception". *Journal of Personality and Social Psychology* Vol. 32, No.5: 880-892.

Schacter, Stanley and Jerome Singer (1962). "Cognitive, Social, and Physiological Determinants of Emotional States." *Psychological Review* Vol.69, No.5: 379-399.

Shapiro, Barbara *et al.* (1981). "Mechanisms of Confabulation". *Neurology* 31: 1070-1076.

Skinner, B.F. (1948) "'Superstition' in the Pigeon." *Journal of Experimental Psychology* 38: 168-172.

Skinner, B.F. (1957). Verbal Behaviour. Cambridge: Prentice- Hall.

Sober, Elliot (1994). "No Model, No Inference: A Bayesian Primer on the Grue Problem". in D. Stalker (Ed.) Grue! The New Riddle of Induction. Chicago: Open Court, 225-240.

Stuss, Donald T. *et al.* (1978). "An Extraordinary Form of Confabulation." *Neurology* 28: 1166-1172.

Snyder, Mark and William Swann (1978). "Behavioural Confirmation in Social Interaction". *Journal of Experimental Social Psychology* 14: 148-162.

Tye, Michael (1997). "Is there Anything it is Like to be a Honey Bee?" *Philosophical Studies* 88: 289-317 van Fraassen, B. (1980). The Scientific Image. Oxford: Clarendon.

Vygotsky, Lev (1934). Thought and Language. Ed., Trans. by Alex Kozulin. Cambridge: MIT Press, 1986.

Wason, P.C. (1966). "Reasoning". in Foss, B.M. (Ed.). New Horizons in Psychology. Harmondworth: Penguin.

Wason, P.C. (1960). "On the Failure to Eliminate Hypotheses in a Conceptual Task". *Quarterly Journal of Experimental Psychology* 12: 129-140.

Wason, P.C. and P.N. Johnson-Laird (1972). Psychology of Reasoning. London: Batsford.

Weinstein, Edwin *et al.* (1956). "Confabulation as a Social Process". *Psychiatry* 19: 383-396.

Wilson, Timothy and Richard Nisbett (1977). "The Accuracy of Verbal Reports About the Effects of Stimuli on Evaluations and Behaviour." *Social Psychology* Vol.41, No.2: 118-131.