Input-Expectation Discrepancy Reduction: A Ubiquitous Mechanism

Derek Partridge
Computing Research Laboratory

New Mexico State University
Las Cruces, NM. USA 88003
phone number: (505) 646-3723

ABSTRACT

This paper examines the various manifestations of input-expectation discrepancy that occurs in a broad spectrum of research on intelligent behavior. It makes the point that each of the different research activities highlights different aspects of an input-expectation reduction mechanism and neglects others.

A comprehensive view of this mechanism has been constructed and applied in the design of a cognitive industrial robot. The mechanism is explained as both a key for machine learning strategies, and a guide for the selection of appropriate memory structures to support intelligent behavior.

A. Introduction

This paper is an attempt to integrate and unify a spectrum of theories about intelligent behavior. I will make the claim that input-expectation discrepancy reduction is the crux of a generic strategy that underlies intelligent behavior. I then go one step further and argue that this mechanism may itself be a particular example of the even more general strategy of primary drive reduction or fitness enhancement. Thus we arrive at a mechanism that derives support from the basic mechanism of life — organic evolution.

I attempt to draw together work from cognitive modelling, experimental psychology, AI, and brain modelling. The motivations behind AI work are seldom this broad, perhaps because AI researchers often have an overly parochial view of what is directly relevant and thus of interest. I hope to demonstrate that each of these viewpoints carries with it a set of biases — certain aspects of the phenomenon are emphasized and others are neglected. Thus, it is by surveying such a range of approaches that we can hope to construct a reasonably complete and unbiased view of the mechanism of interest.

B. Scenes, Scripts, Plans, MOPs, TOPs, etc.

Taking a top-down approach, we find Schank (1982) advocating that high-level cognitive activity (such as engaging in day-to-day dialogue) is mediated by a complex set of interrelated memory structures: TOPs (Thematic Organization Packets), MOPs (Memory Organization Packets), Scripts, Scenes, and memories. A major part of Schank's thesis is that learning is driven by expectation failures from predictions encoded in memory.

For Schank, "the dominant notion in building and altering memory structures is expectation-failure." When this happens he offers three general possibilities:

(A) Modify specific expectation

(B) Alter script itself

(C) Index as expectation failure

These three classes of memory modification correspond to fine tuning, a generalization, and expecting an accepted anomaly in the generalized information, respectively.

A first time failure is indexed as an exceptional occurrence. Repetition of similar failures suggests that what we believed was exceptional is perhaps quite normal and so a more drastic revision of memory structures is called for: either replacement of an entire structure, or reorganization of the placement of a structure. The decision between these two alternatives is, according to Schank, based upon the degree of overall success that we have had with this structure in the past. If it has in general worked well then we keep it and reorganize; if it has not worked well then we will replace it. Here we see a first appearance of a 'confidence' measure associated with the learning process, other studies elaborate on this aspect of the general mechanism.

Notice also that expectation failure is the cue for modifying expectations, no mention is made of modifying the perceptual mechanisms. In the language of a study discussed below, the input and expectation disagree — perhaps we misperceived the input? Schank doesn't seem to address this possibility.

It is true that in typical Schankian contexts (i.e., restaurants) it is difficult to believe that you could "perceive" that you paid before the food arrived when actually you paid as normal, after eating and just before leaving. But in other contexts (e.g. natural language communication) gross perceptual errors are quite possible. The possibility for error in the processing and interpretation of sensory information adds more complexity to the problem of expectation failure and subsequent learning.

Schank's term "expectation failure" proclaims this asymmetry - when sensory information and expectation don't agree, it is the expectation that is deemed to have failed. In order to emphasize that there are two sides to this lack of agreement (as in all quarrels) I prefer the term "input-expectation (i - *e)* discrepancy."

To the extent that expectations may also drive top-down perceptual processing, misperception may also be expectation failure. There are two points here: first, by no means is all perception top-down. Second, it may still be useful to maintain a separation between

the processing of sensory information and the subsequent comparison with memory structures.

The extent to which *i - e* discrepancy is due to the perceived input (I take i ''s to be interpretations of raw data — in the head or system rather than the world) rather than the expectation being wrong will depend, in general, upon both the unfamiliarity of the perceiver with a particular context, and the extent to which the context is assumed to be strange. Thus perceptual errors will be most likely when the perceiver is confident that he is in a familiar context when in fact he isn't.

A literary work that always makes this point forcibly to me is the Alexandria Quartet by Lawrence Durrell. The first three volumes treat the same general sequence of events from the perspectives of three different people. Each of the three perceives the sequence of events in a self-consistent, quite believable manner (given Durrell's penchant for the fantastic), but each perceives certain critical events totally differently and therefore has a radically different explanation of what they witnessed. In fact what we are treated to in these books is expectation confirmation, there was no f - *e* discrepancy despite the fact that there were three very different e 's and only one i on a number of occasions. The point is of course, there were also three different *i* 's after perceptual processing, and what is more, each individual's expectations influenced the *i* 's that they perceived.

Inputs and expectations are not, in general, independent of each other and the dependence works both ways. A last point is that because of this interdependence the occurrence of an expectation failure or an i - *t* discrepancy depends upon some higher level control: we can refuse to admit that discrepancies exist, or insist on their existence as dictated by some higher level goals.

Thus to borrow Schank's favorite situation, the restaurant: if I eat at a particularly expensive restaurant and the food is poor, I might well resist acknowledging the failure of my expectation that the food will be good because I wish to preserve my general belief that I always spend money wisely. This view of i - *e* discrepancy takes us into the realm of "dissonance theory" (see Aronson, 1078) and falls into the category of inconsistencies between one cognition and a more general, more encompassing cognition. Dissonance theory suggests that individuals will strive to reduce such dissonance, for example, by refusing to acknowledge the poor quality of the expensive food.

Rumelhart and Ortony (1977) do emphasize top-down processing which leads "from conceptual expectations towards the data in the input where satisfaction of these expectations might be found." They suggest that finding a good fit between expectations and input (i.e., minimizing i - e discrepancy) is a critical part of the strategy for selection of appropriate memory structures (schemata) from the enormous number of possible schemata — a context-directed selection process.

From this perspective, expectation failure (or more accurately, i - e discrepancy) is not the cue for learning, but for eliminating the schemata responsible for the failed expectations from the set of potentially appropriate schemata for comprehension of the current situation. This is, of course, not the function of expectation failure in Schank's model at all.

I am not suggesting that these two theories are contradictory, only that their advocates are emphasizing different aspects of a very complex process. Human information processing is a non-trivial combination of top-down and bottom-up mechanisms, and recognition and understanding are not separate processes.

Rumelhart and Ortony summarize the processes: "information (including both the stimulus and the context) enters the system and directly suggests certain plausible candidate schemata to account for it. At the same time as this data driven processing is going on, such postulated schemata activate their dominating schemata, which in turn look for other as yet unsuspected aspects of the situation... A schema is said to provide a good account of (aspects of) the input situation when it can find good evidence for itself."

The above theorizing falls into the class of the "theory development methodology" in cognitive science (Miller, 1978). A sufficient explanation of cognitive phenomena is being sought without undue concern for the existence of empirical consequences of the theory.

C. Lexical Decisions, RT and Mismatch Detectors, etc.

The second class of investigation of an *i - e* discrepancy reduction mechanism favors the term 'mismatch detector' and epitomizes the alternative to the "theory development methodology" -- it is "theory demonstration methodology." This approach to theorizing demands empirical testability of a proposed theory, and consequently, the theorizing is limited to highly controlled and thus somewhat artificial phenomena.

The evidence for mismatch detectors has been sought in experimental paradigms that involve word comprehension. Lexical decision tasks, such as word or nonword discriminations applied to a target string of letters, constitute one source of empirical evidence for and against theories of language understanding.

The "verification model" (Becker, Schvanevcldt, and Gomez, 1973, and Becker, 1980) attempts to account for the context effects observed in lexical decision tasks (word or nonword response to a target stimulus that follows a cue stimulus). This model postulates mechanisms that generate two sets of expectations: the sensory set (generated on the basis of sensory features extracted from the target stimulus — i.e., structural or 'syntactic' similarity), and the semantic set (generated on the basis of a semantic similarity to the cue stimulus). The semantic set is searched first during the verification process which attempts to effect recognition of the target stimulus.

Becker (1980) postulates a "prediction" strategy (when there is a small semantic set size due to the cue-target pairs being highly related), which we can view as involving a focused expectation, and an "expectancy" strategy (when semantic set size is large), which generates a broader, unfocused expectation. This class of research emphasizes the use of mismatch or *i - e* discrepancy as a guide to the selection of correct perceptions in the style of Rumelhart and Ortony, and in sharp contrast to Schank — learning behavior is neglected.

Becker (1980) states that it is a common assumption (although not one that goes unchallenged) that the processes isolated in word recognition are the same as those involved in fluent reading skills. His final suggestion is that the types of strategies he describes are

indeed general strategies and thus we should detect their operation in tasks involving, say, the perception of pictures. An event may be divided into a sequence of snapshots. We might then take a pair of snapshots and present them as a cue-target pair. Thus in the context of a restaurant, a pair like "order meal/eat meal" might lead to the use of the "prediction" strategy, and "order meal/pay for meal" might yield "expectancy" strategy effects.

Now the word recognition task paradigm has clearly been stretched up into Schank's domain, in more than one sense. But we are, in general, on the way down seeking the more reductionists uses of i - c discrepancy. Staying with approximately this level of phenomena (i.e., word recognition and more generally, reading) we can leave the jungle of RTs and %-correct, and examine empirical data that is directly tied to the observable mechanisms of the brain.

D. Along Come Brain Waves

The use of computer analysis has enabled the isolation of stimulus-locked segments of the electrical activity of the brain. The electrical activity recorded during and shortly after the presentation or expected presentation of a stimulus is called an event related potential (ERP). Extensive research has shown that certain components of the ERP are sensitive to a person's expectations. In particular, unexpected or novel stimuli are typically followed after some 300 msec, by a positive ERP component known as the P_3, which has also been shown to be well-correlated with other indices of orienting (e.g. pupil dilation).

The subjective probability (or confidence in the expectation) of a stimulus and the value, utility, or relevance or a stimulus are two classes of variables that appear to affect the amplitude of the P_3 component. In general, P_3 amplitude increases with both the unexpectedness and the value of a stimulus (Johnston, 1979). P_3 has been shown to be influenced by a number of other variables.

For my current purposes, P_3 appears to be an electrophysiological indication of i - *e* discrepancy or expectation failure. Despite the wealth of ERP experiments it is only recently that linguistic material has been used in ERP tasks. Kutas and Hillyard (1980) have investigated ERPs in the context of a sentence reading task. They state that the language comprehension task has often been characterized as a continuous testing and updating of hypotheses about the words that are likely to occur next in a text or conversation. They found that semantically inappropriate words (i.e., "He spread the warm bread with *socks.")* elicited a late negative wave (N400). This wave may be, they argue, an electrophysiological sign of the "reprocessing" of semantically anomalous information — the result of an *i - e* discrepancy, in this case, expectation failure. Fainsilber, Miller, and Ortony (1984) examined N200/400 to assess whether the detection of anomaly is an integral part of understanding metaphor, and concluded that "understanding a metaphorical comparison appears to involve the registration of a mismatch, whereas understanding of a literal comparison does not."

Holcomb (1983) notes that the N400 findings fit well with Becker's verification model: the "N400 component in many ways appears to resemble a semantic mismatch detector. N400 is large when the probability of another word occurring is great, but only if the expectation was based on semantic information. In the present framework, an ERP model of semantic context effects, N400 is proposed to represent the activity of an automatic semantic mismatch detector."

E. Down to the Neurons

Moving on down to a more reductionist view of t - e discrepancy, Partridge, Johnston, and Lopez (Johnston et al., 1983, Partridge et al., 1984), have theorized, modelled, and presented results of a detailed physiological mechanism for generating expectations and modifying memory structures as a result of expectation failure.

The theorizing was based on a number of sources: Sokolov (1963) suggested that neural 'models' of our expectations are constructed and modified as a result of the "impulses of discrepancy" encountered (an early parallel of Schank's suggestions, one that is tied to relatively concrete representational structures but lacks the depth of Schank theories); cell assembly theory (originated by Hebb, 1949) gives us physiologically-based units for distributing and maintaining 'activity' in a neural network; and the existence of empirical data relating the magnitude of the orienting response (OR) to variables underlying unexpected stimuli.

In addition, the basic learning behaviors accounted for (described below) are sufficiently low-level and ubiquitous in the animal world that we can postulate *why* they might exist in terms of evolution theory and survival. Expectation failure or i - e discrepancy from this biological perspective can be viewed as a specific class of more general mechanisms: a genetically determined motivation to satisfy primary drives (or goals') — such as find food and avoid pain. Stated somewhat simplistically, the importance of satisfying these goals is that they are a basis for survival, and survival is a key idea in the theory of evolution. Hence we might reasonably expect the products of evolution to be genetically preprogrammed with efficient mechanisms for satisfying these goals.

The final major hypothesis is that the efficient selection and assimilation of useful information is also a fundamental survival goal analogous to the more conventional ones. The key to the selection mechanism is the unexpectedness or novelty of the information. Given the uncertainty of the empirical world and an organism that attempts to predict its future (the better predictors will be the survivors), the predictions will sometimes, and to some extent, fail — this mismatch, failure, or discrepancy is the key to the selection mechanism. Thus we postulate a novelty drive mechanism, which is just another way of saying mismatch detector or expectation failure mechanism, except that it implies that the mechanism will be analogous to the other basic drive mechanisms such as hunger drive. Hence, theories and data pertaining to these conventional drive mechanisms should provide insight into the mechanism of t - *e* discrepancy reduction and human learning. So we find yet another potential source of information for elucidating human learning mechanisms.

Rescorla and Holland (1976) divided basic learning behaviors into three categories:

(a) single stimulus presentations;

(b) exposure to relations among stimuli; and

(c) exposure to relations between responses and

stimuli.

Johnston, Partridge, and Lopez (1983) modelled the novelty drive theory and demonstrated that it accounted for a wide range of empirical data on human learning in categories (a) and (b) above.

A cell assembly network generates an expectation of the next stimulus input that it will encounter. Any i - *e* discrepancy (that exceeds certain thresholds - exact matches are not a feature of reality) is interpreted as an expectation failure (the environment is scanned and an internal representation of the input stimulus is generated but there was no provision for questioning the t 'perceived').

It is argued (Johnston et al., 1983, Partridge et al., 1984), that the novelty or unexpectedness of an input with respect to an expectation is:

/ (qualitative difference between i and c ;

quantitative difference between i and e)

and that OR is:

/ ' (novelty; *i* - *e* magnitude)

where t - e magnitude is some "additive" function of magnitude of the expected stimulus and magnitude of the actual input stimulus.

Apart from suggesting and allowing the exploration of such details of the mechanism, the study also demonstrated that the hypothesis that the *i* - *t* discrepancy reduction mechanism can be structured as a conventional drive mechanism (i.e., novelty drive) is a viable one. An implication of this hypothesis is that the unexpectedness of a stimulus is a primary drive reducing quality just like several others, such as food, sex, and pain. Whereas Schank, for example, states that "When we see something that in no way surprises us, it is also likely that it in no way interests us either."' However, the importance of an input stimulus is not just its unexpectedness, but more fundamentally, the importance is the potential for reducing a primary drive. In AI terms, this importance is the potential for achieving some basic goal, such as a need for novelty when bored, a need for food when hungry, etc. As mentioned earlier some ERP data is supportive of this view — the magnitude of the OR appears to be a function of both unexpectedness and utility (i.e., food is an arousing stimulus even if expected when we are hungry).

The basic learning behavior — habituation (learning not to respond), may sound fairly trivial to the AI researcher, but it is in fact crucial at all levels of behavior. An organism must ignore most of the information that bombards its sensory organs - it is an apparent efficiency measure that the real world promotes to one of necessity.

Consider the AI paradigm of rule-learning but in the empirical world rather than an abstracted context characterized by drastically pruned descriptions. The act of describing removes most (perhaps all) of the potentially relevant, but actually irrelevant, attributes of each event before the learning algorithm sets to work — it is fed predigested reality, hence no need to learn when and what to ignore. The child that is always fed filleted fish is not very impressed by a technique for avoiding bones.

Next I shall describe a mainstream AI application that makes important use of the *i* - *t* discrepancy reduction mechanism and can be used to illustrate many of the specific biases inherent in each of the above-described approaches to this mechanism.

F. A Cognitive Industrial Robot

High level control mechanisms for industrial robotics applications have been designed, implemented, and partially tested (Partridge, Burleson, and Lopez, 1985) — the hand-eve robot learns and reasons (hence a "cognitive" robot) about a general task plan with respect to a specific task setup and attempts to optimize its performance. The mechanism behind the attempted optimization is to learn the constancies in a flexibly fixtured environment (e.g. that a certain target object tends to be situated in a certain position). This learned knowledge is then used predictively to optimize task execution in the current environmental setup.

The context of industrial robotics provides a constrained and thus potentially tractable micro-world but one that also contains much potential for the effective application of fundamental AI techniques. There is potential for developing a rich knowledge structure (certainly richer than the cell assembly model but not as rich as that of Schank's theories) within a set of constraints that both promise tractability, and yet still offer a realistic and thus potentially testable context (in contrast to both Schank's theories which are not currently very testable, and the RT paradigms which are testable but highly artificial). A cognitive industrial robot can fill a real need and offers some hope of comparison with the human performance of similar tasks.

Of major importance for the current discussion is the implementation of the *i* - c discrepancy reduction mechanism within this project. Having learned that some significant object (a target object with respect to this sub-task), say object *A* , has tended to be situated at position (x, y) whenever it was required, the robot might predict the future occurrence of *A* at (x, y). The expectation (in the context of this particular sub-task) that *A* will be at (x, y) can be used to drive a top-down pattern recognition process — i.e., the process just checks that *A* is at (x, y) rather than analyzing the input image bottom-up to find *A* . When *A* is at (x, y) the robot confirms its expectations quickly and proceeds to deal with object *A* as dictated by the current sub-task (e.g. it might pick it up).

But when the input and the expectation don't agree (object *A* does not appear to be at (x, y)), then the roDOt needs to learn something — the big question is: what?

According to Schank's theory the expectation that object *A* would be at (x, y) failed, and it needs to change its memory structures. On first failure it just indexes the general expectations with the exceptional possibility that object *A* may alternatively be at (x_1, y_1) — the position where it was eventually found as a result of bottom-up recognition. On subsequent failures of this expectation, the robot should learn that its general expectation of *A* at (x, y) no longer holds and should be abandoned. It should be replaced by the expectation of *A* at (x_1, y_1) if this hitherto exceptional possibility has been repeatedly encountered. Alternatively, the general expectation that *A* will be situated in any particular position may be completely dropped if the failures were due to seemingly arbitrary sequences of positionings of object *A* - an environmental constancy has disappeared, or was erroneously learned in the first place, in either case it can no longer be

exploited.

A second possibility once we have decided that expectation failure is the source of the i - e discrepancy is that the expectation itself may be correct but the context that it was generated from was wrong. This is perhaps just to say that the subsequent learning should be at a higher level, i.e., the structure that selects the context of the current stimuli should be altered so that in the future it will select the correct context. Schank does raise this type of question, that of level of learning, and he sketches out some answers.

But as mentioned earlier, Rumelhart and Ortony suggest that such expectation failure may be a crucial factor in the correct selection of an appropriate context and not a learning situation at all. They offer the view that minimizing i' - *e* discrepancy is the route to efficient selection of an appropriate context. The verification model also uses expectation failure as a selection mechanism.

Similarly, the cognitive robot may use this best fit between input and expectation to efficiently select the appropriate context after a discontinuity in task execution due to the occurrence of an error condition. As part of the analysis of the error condition it might need to choose between, say, an expectation of *A* at *(x, y)* or *B* at *(x_1 y_1)*.

But in this robotics context, an i - e discrepancy can be due to a problem with the i component. First, the input image may be of poor quality due to 'noisy' conditions, in which case it might be appropriate to question the raw input data and our interpretation of it that yielded the *i* that was discrepant.

For the cognitive robot expecting object *A* at position *(x, y)*, a subsequent *i - e* discrepancy may be due to the fact that although *A* was indeed at *(x, y)* the input pattern was so degraded by 'noise' that the cursory top-down recognition algorithm failed to confirm its expected presence. A subsequent, more exhaustive, bottom-up analysis might well find *A* at *(x, y)* with a sufficiently high confidence (it might for instance have the information that *A* is definitely somewhere in the image, and it might determine that the patterns at all other positions resemble object *A* even less that the pattern at *(x, y)*).

The problem with the i component may be due to the *interpretation* of the raw data rather than the quality of the data itself. This possibility then leads us into the murky world of i - e interdependencies: interpretation of the raw data depends upon expectancies, and expectancies depend upon interpretations of the data.

The lexical decision tasks described earlier have been used to probe the complexities of this *i - e* interdependence. Schvaneveldt and McDonald (1981) report on a series of six such experiments that were designed to investigate the role of semantic context in the perceptual process. Their results support the view that there are two modes of processing sensory information.

One mode involves the initial analysis of sensory information (such as features of the stimulus or such holistic properties as word shape) and is not directly affected by semantic context. The second mode is characterized as a "second look" at the stimulus (remember the N400 component of the ERP?). They view this secondary analysis as "basically a memory-driven process in which 'hypotheses' about the identity of the stimulus are tested by comparing actual stimulus characteristics with those predicted by the hypothesis. The hypotheses are generated by a combination of sensory information (from the first mode of processing) and contextual information." Thus an initial, partial t is generated independent of any *e* , this t is then combined with contextual information to yield e 's that guide the final step in the recognition process.

In terms of the popular general theory that the analysis *of sensory* information is compounded *of* a succession of processes which depend to varying degrees on bottom-up and top-down modes of organization, Schvaneveldt and McDonald suggest that the initial process is independent of expectations while the later ones are directed by expectations. They also suggest that the top-down process may enhance perception of discrepancies rather than induce a perceptual or decision bias in favor of expected stimuli.

Subsequently, Paap, Newsome, McDonald and Schvaneveldt (1982) described a development of the verification model, called the "activation-verification" model. Their goal is to specify the nature and interaction of bottom-up and top-down information-processing activities in recognition. The solution provides an independent top-down process (verification) that involves comparing stimulus information to prototypes stored in memory.

The suggestion that a top-down process may enhance perception of discrepancies raises again the basic problem of whether or not an i - e discrepancy exists in any particular situation. Discrepancies may be perceived or unperceived dependent upon both the general level and the detailed focus of an *i - e* comparison. A high-level comparison will eliminate low-level discrepancies which may well be appropriate in the empirical world where repetition is never exact. On the other hand, some details of a stimulus are likely to be important while others are not, hence the *% - e* comparison needs to be focused on the significant details only. (Also a fundamental problem in machine learning: how are the significant features in a series of events selected?)

There are two general classes of misinterpretation of the input stimulus:

(1) misinterpretation due to erroneous recognition of input data, e.g. an input pattern generated by object *A* may be erroneously recognized as object *B*; and

(2) misinterpretation due to erroneous selection from the input data, e.g. a pattern generated by noise is recognized as object *A* whilst the pattern generated by object *A* is dismissed as noise.

In the cognitive industrial robot such misinterpretations are more likely when recognition is top-down (hypothesis or expectation driven), and when the system is 'confident' in its expectation.

This raises the last feature of the *i - e* discrepancy reduction mechanism: confidence. Confidences in both the i 's and the *t* 's obtained interact with general contextual confidences to influence the following:

(a) whether or not we perceive an i - *e* discrepancy; and

(b) how to analyze a perceived discrepancy.

Problem (b) has largely been dealt with above except to note that analysis of the cause of an *i - e* discrepancy can be guided by the confidence that the system has in

the correctness of the structures and processes that are applied. The obvious strategy is to suspect that the element in which we have least confidence is the most likely cause of the discrepancy perceived.

Apart from the fact that the above rule is just a heuristic and as such carries no guarantees (in addition it has not, to my knowledge, been tested in this context although the cognitive robot uses it). There is, in addition, no correct way to compute these confidences — again we must explore heuristics until we find an adequate strategy. Another common AI problem: confidence ratings (or some similarly named attribute) appear to be necessary but there is no obvious and correct way to compute them.

The cognitive industrial robot generates a confidence in its expectation that, say, object *A* is at *(x, y)* based upon the relative frequency of past occurrence of *A* at *(x, y).* It also generates a confidence that *A* is actually at *(x, y)*. This confidence is based upon both the degree to which the actual features of the pattern found at *(x, y)* have been matched against the characteristic features of *A* (the number of features and how well they matched), and the degree to which recognition has been top-down rather than bottom-up (a cursory top-down analysis is more efficient but less reliable).

Problem (a) concerns the conditions under which we acknowledge the presence of an i' - *e* discrepancy. Very roughly, the more confident we are that we understand the situation, the less likely we are to admit that there is a discrepancy.

But, you might object, either i and *e* match or they don't — there should be no question of confidence here. In an idealized situation this may be true, but the real world is far from ideal. Two aspects of reality suggest the necessity for confidence ratings:

(i) Exactly the same event never occurs twice (it is only abstractions from the sensory data that exactly repeat), thus even the best i - *e* match will not be perfect; there will always be some discrepancy.

(ii) Stimuli are not just given, they must be selected from a rich and complex continuum; variation in the selection of appropriate attributes will result in varying discrepancies.

As mentioned earlier, point (ii) is a well-known and difficult AI problem. One approach to this prob-0 lem is through the goals of the system. Thus the cognitive robot, for example, has a major goal of increasing task efficiency, or reducing task execution time; hence time is an important attribute of any subtask and one that it must always select. But, in general, selection of the significant aspects of its environment is subtask dependent.

Problem (i) suggests that the mechanism is not founded on an *i - e* discrepancy itself, but on a discrepancy that exceeds some threshold. Compounded with this there is not one threshold but one for each significant attribute of the stimulus. Finally these thresholds are dynamically adjustable and a major factor in this dynamic adjustment is confidence - if confidence is high then the thresholds, in general, are raised. High confidence suggests that we will ignore larger discrepancies.

G. Summary

A range of projects has been surveyed in an attempt to demonstrate the potential utility to AI of work that lies outside the normal concerns of mainstream AI researchers. In particular, a generic mechanism that appears to underlie intelligent behavior was examined — the t - e discrepancy reduction mechanism. This mechanism appears to play a key role in human learning and in control of cognition. It is thus expected to be of importance in AI both as a key to machine learning (when to learn and what to learn - two major unsolved AI problems), and in the control of complex context selection.

It was shown that a comprehensive understanding of this mechanism is obtained from a consideration of this range of approaches to it; each approach embodies a different set of biases. First, there are two general applications of the *i - e* discrepancy reduction mechanism:

(i) as the basis for a selection mechanism — selection of 'best-fit' contexts at one level, and of words at another level (it is a focusing mechanism); and

(ii) as the cue for a learning mechanism.

Within the latter application the presence of an i - e discrepancy signals the need for learning, but different research has emphasized different details of interpretation of this discrepancy to guide what needs to be learned. Did the expectation fail? Or the interpretation of the input? Or both? Or, from a higher level viewpoint, was some aspect of the i - e comparison itself misconceived? This last possibility leads us back to the first general application above - points (i) and (ii) are not independent.

It was further argued that *i - e* discrepancy reduction might itself be a special case of the more general mechanism of basic goal achievement — the fundamental mechanisms of survival.

H. Acknowledgements

I wish to thank Victor Johnston, Patty Lopez, Andrew Ortony, and Roger Schvaneveldt for their assistance and guidance in my forays into domains that I am only just beginning to appreciate.

1. References

[1] Aronson, E. (1979) The Theory of Cognitive Dissonance: A Current Perspective, in *Cognitive Theories in Social Psychology,* L. Berkowitz (ed.), Academic Press: NY, pp. 181-220.

[2] Becker, C.A. (1980) Semantic context effects in visual word recognition: An analysis of semantic strategies, *Memory and Cognition,* 2, 6, pp. 493-512.

[3] Becker. C.A., Schvaneveldt, R.W., and Gomez, L.M. (1973) Semantic, Graphemic, and Phonetic Factors in Word Recognition, *Psychonomics Society,* St. Louis, Missouri.

[4] Fainsiiber, L., Miller, G.Z., and Ortony, A. (1984) An Electrophysiological Study of Metaphor Comprehension, *Society for Psychophysiological Research,* Milwaukee, Wisconsin.

[5] Hebb, D.O. (1949) *The Organization of Behavior,* Wiley: NY.

[6] Holcomb, P.J. (1983) Automatic and strategic attention: An event-related brain potential analysis of contextual processing, *Ph.D. Dissertation,* New Mexico State University.

[7] Johnston, V.S. (1979) Stimuli with Biological Significance, in *Evoked Brain Potentials and Behavior,* H. Begleiter (ed.), Plenum Press: NY, pp. 1-12.

[8] Johnston, V.S., Partridge, D., and Lopez,P.D. A Neural Theory of Cognitive Development, *Journal of Theoretical Biology,* 100, pp. 485-509, 1983.

[9] Kutas, M. and Hillyard, S.A. (1980) Reading senseless sentences, *Science,* 207, pp. 203-205.

[10] Miller, L. (1978) Has Artificial Intelligence Contributed to an Understanding of the Human Mind? A Critique of Arguments For and Against, *Cognitive Science,* 2, pp. 111-127.

[II] Paap, K.R., Newsome, S.L., McDonald, J.E., and Schvane veldt, R.W. (1982) An activation-verification model for letter and word recognition: the word-superiority effect, *Psych. Rev.,* 89, 5, pp.573-594, 1982.

[12] Partridge, D. Johnston, V.S., and Lopez, P.D. (1983) Computer Programs as Theories in Biology, *Journal of Theoretical Biology,* 108, pp. 539-564.

[13] Partridge, D., Johnston, V.S., Lopez, P.D., and Burleson, C. (1985) A Hand-Eye Coordination Algorithm For A Cognitive Industrial Robot, *procs. 18th Systems Science Conference,* Hawaii.

[14] Rescorla, R.A. and Holland, P.C. (1976) . Sonie behavioral approaches to the study of learning, in *Neural Mechanisms of Learning and Memory,* M.R. Rosenweig and E.L. Bennett (eds.), MIT Press: Cambridge, MA, pp. 165-192.

[15] Rumelhart, D.E., and Ortony, A. (1977) The representation of knowledge in memory, in *Schooling and the Acquisition of Knowledge,* R.C. Anderson, R.J. Spiro, and W.E. Montague (eds.), Erlbaum: NJ.

[16] Schank, R.C. (1982) *Dynamic Memory,* Cambridge University Press: Cambridge, U.K.

[17] Schvaneveldt, R.W. and McDonald, J.E. (1981) Semantic context and the encoding of words: evidence for two modes of stimulus analysis, J. *Exp. Psych., 7,* 3, pp. 673-687.

[18] Sokolov, E.N. (1963) *Perception and the Conditioned Reflex,* Pergammon Press: NY.