Autism and Coherence: A Computational Model

CLAIRE O'LOUGHLIN AND PAUL THAGARD

Abstract: Recent theorizing about the nature of the cognitive impairment in autism suggests that autistic individuals display abnormally weak central coherence, the capacity to integrate information in order to make sense of one's environment. Our article shows the relevance of computational models of coherence to the understanding of weak central coherence. Using a theory of coherence as constraint satisfaction, we show how weak coherence can be simulated in a connectionist network that has unusually high inhibition compared to excitation. This connectionist model simulates autistic behaviour on both the false belief task and the homograph task.

1. Introduction

In her influential monograph *Autism: Explaining the Enigma*, Uta Frith (1989) presents a parsimonious account of the complex mix of problems seen in autism. Her account identifies an abnormal pattern in the processing of information by autistic subjects: across a wide range of cognitive tasks there is a surprising failure to process meaningful and patterned stimuli more effectively than stimuli that are random and devoid of structure (Frith, 1970a, 1970b; Hermelin and O'Connor, 1970). On the basis of this pattern, Frith (1989) postulates that autism is marked by a reduced capacity to integrate information at different levels. This proposal rests on two claims: firstly, that our cognitive apparatus is predisposed toward the synthesis of incoming information from the environment as a way of deriving meaningful and coherent experiences; and secondly, that this capacity for coherence-based inference is substantially weakened in the autistic case. The result according to Frith is a peculiar processing style, characterized by a tendency to deal with information on a piecemeal basis, conjoined with a relative inability to situate and interpret information within a wider relevant context.

The hypothesis that individuals with autism demonstrate piecemeal rather than integrative processing, has been examined with a range of experimental materials noted for their gestalt-inducing qualities, where weak central coher-

Claire O'Loughlin was supported by a Claude McCarthy Fellowship, and Paul Thagard was supported by the Natural Sciences and Engineering Research Council of Canada and by a Canada Council Killam Research Fellowship.
We thank Francesca Happé, Michael Dixon, and an anonymous reviewer for helpful comments on a previous draft.
Address for correspondence: Paul Thagard, Department of Philosophy, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.
Email: pthagard@watarts.uwaterloo.ca.

Mind & Language, Vol. 15 No. 4 September 2000, pp. 375–392.

ence would be expected to confer a significant task advantage. For example, Shah and Frith (1993) demonstrated that superior performance by autistic subjects relative to controls on the Wechsler Block Design task was due to a greater ability to segment the whole design into its component parts. Similarly, Happé (1996) found that individuals with autism were less susceptible than controls to standard visual illusions that depend upon surrounding context for their effect, suggesting that they tended to perceive the figures in a less unified fashion. People with autism have also been found to excel on the Embedded Figures Test (Shah and Frith, 1983) in which a hidden figure must be detected within a larger meaningful picture. By contrast, in a homograph disambiguation task which specifically requires the processing of information in context for its solution, autistic individuals failed to use preceding sentence context to determine the correct pronunciation of the homographs (Frith and Snowling, 1983; Happé, 1997).

Frith's proposal that autistic information processing displays weak central coherence provides a working hypothesis for tackling the major features of autism that have previously defied assimilation under a unified explanatory scheme. She suggests that the noticeably uneven pattern of intellectual abilities encompassing both assets and deficits in performance, the repetitive phenomena of stereotypes and perseverative behaviour, as well as the core impairments in social interaction and communication, are intelligible if viewed as symptoms or manifestations of a more general dysfunction in the capacity for coherence. More recently, Frith and Happé (1994) have reviewed the role of weak coherence in social communication, and suggest that there may be two quite different cognitive abnormalities underlying autism: a specific deficit in a 'theory of mind' module that underwrites social skills, and weak coherence which appears to characterize the processing style of even high functioning autistic individuals. Questions raised by current research (e.g. Happé, 1996, 1997, 1999), include the precise nature of weak central coherence (deficit/cognitive style), how widespread its effects on cognitive functioning are (level/s at which it operates), and the manner in which a weakened capacity for coherence relates to other proposed deficits in 'theory of mind' and executive function (overlap/interaction).

In this paper we deepen the weak coherence theory of autism by providing a model of how coherence deficits can produce the cognitive problems of autistic individuals. First we outline a precise characterization of coherence as constraint satisfaction and describe its implementation in a connectionist network. Then we consider various possibilities for impairing coherence-based inference, and present a model of weak coherence that is capable of reproducing experimental findings with autistic subjects. Finally we discuss implications of our computational model for the ongoing development of this cognitive account of autism.

2. Coherence as Constraint Satisfaction

In order to determine the nature of weak coherence, we need to begin by clarifying the operation of coherence in normal human thought. Frith and Happé (1994) suggest that we are predisposed to pull together pieces of information into coherent patterns, and highlight parallels between this 'predisposition' and the dominance of holistic processing stressed by Gestalt psychologists to reinforce the abnormality of the piecemeal processing style seen in autism. But given that coherence-based inference is taken to be a dominant feature of human cognition, how does it function? When we attempt to make sense of a person or situation, how do we integrate the numerous pieces of available information—taking account of those pieces that fit together and those that do not—to arrive at the most coherent interpretation of that person or situation? And how is it possible to capture the dynamics and complexity inherent in such a process?

While typically underspecified in the philosophical and psychological literature, the concept of coherence has recently been given a precise computational characterization (Thagard and Verbeurgt, 1998; Thagard, forthcoming). They propose that coherence be understood in terms of maximal satisfaction of multiple competing constraints. An informal summary of their theory of coherence can be given as follows:

(1) Elements are representations such as concepts, propositions, parts of images, goals, actions etc.
(2) Elements can cohere (fit together) or incohere (resist fitting together). Coherence relations include explanation, deduction, facilitation, and association. Incoherence relations include inconsistency, incompatibility, and negative association.
(3) If two elements cohere, there is a positive constraint between them. If two elements incohere, there is a negative constraint between them.
(4) Elements are to be divided into ones that are accepted and ones that are rejected.
(5) A positive constraint between two elements can be satisfied either by accepting both of the elements or by rejecting both of the elements.
(6) A negative constraint between two elements can be satisfied only by accepting one element and rejecting the other.
(7) The coherence problem consists of dividing a set of elements into accepted and rejected sets in a way that satisfies the most constraints.

More precisely, consider a set E of elements that may be propositions or other representations. Two members of E, e1 and e2 may cohere with each other because of some relation between them, or they may resist cohering with each other because of some other relation. Making E into as coherent a

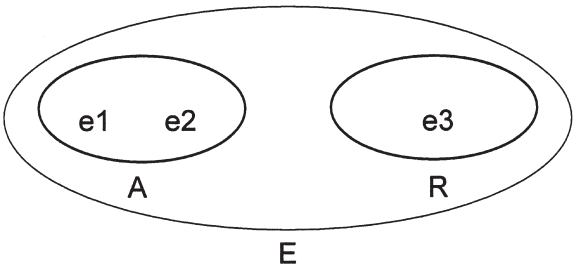

Figure 1

whole as possible involves taking into account the coherence and incoherence relations that hold between pairs of members of E. This is achieved by partitioning E into two disjoint subsets, A and R, where A contains the accepted elements of E, and R contains the rejected elements of E. This partition is performed in such a way so as to maximize compliance with the following two coherence conditions:

(1) if two elements (e1,e2) are positively constrained, then e1 is in A if and only if e2 is in A.
(2) if two elements (e1,e2) are negatively constrained, then e1 is in A if and only if e2 is in R.

For example, if E is a set of propositions and e1 explains e2, we want to ensure that if e1 is accepted into A, then so is e2. On the other hand, if e1 is inconsistent with e3, we want to ensure that if e1 is accepted into A, then e3 is rejected into R (Figure 1). The relations of explanation and inconsistency provide constraints on how we decide what can be accepted and rejected.

Adopting this characterization of coherence enables many kinds of cognition to be meaningfully analysed as coherence problems. Specifically it demonstrates how we are able to integrate numerous pieces of information in order to make sense of our physical and social surroundings; that is, how we move from pieces of information that are coherent/incoherent with each other at a local level to a global interpretation of coherence. More generally, characterizing coherence in terms of maximal constraint satisfaction recognizes that in many real-life situations it will not be possible to satisfy all the relevant constraints operating, but that nevertheless we generally strive to satisfy as many as possible. For example, when interpreting someone's behaviour, we want an interpretation that makes the most sense of the available information (as opposed to any interpretation). The idea of coherence maximization captures this intuition.

In order to apply this general account of coherence as constraint satisfaction to particular psychological phenomena, it is further necessary to specify the types of elements and constraints involved, and provide an appropriate

interpretation of acceptance and rejection for each kind of coherence problem. Thagard and Verbeurgt (1998) summarize a range of applications of this characterization of coherence to processes as varied as explanatory inference (Thagard, 1992), decision making (Thagard and Millgram, 1995), analogical mapping (Holyoak and Thagard, 1995), and interpersonal impression formation (Kunda and Thagard, 1996).

3. Computing Coherence

Coherence construed as maximizing the satisfaction of positive and negative constraints among elements is naturally computed using connectionist (neural network) models. In a connectionist network, the elements involved in the coherence problem are represented by nodes or neuron-like units. Positive constraints are represented by excitatory connections and negative constraints by inhibitory connections. Hence, if two elements are positively constrained, then the nodes representing these elements are connected by an excitatory link. Conversely, if two elements are negatively constrained, then the nodes representing these elements are connected by an inhibitory link.

Once the constraint network has been constructed and all the nodes assigned an initial activation, coherence can be computed by a process that involves repeated cycles of activation adjustment. In each cycle the activation of all the nodes is adjusted using a parallel updating algorithm, with each node being updated on the basis of the activation of the nodes to which it is connected by excitatory and inhibitory links. This process is repeated until all of the nodes have reached stable activation levels and the network is referred to as having settled. The final activation levels of the nodes determine which elements will be accepted and which will be rejected. Elements represented by nodes that have a final activation above a specified threshold (usually 0) are accepted, and elements represented by nodes that have a final activation below this threshold are rejected. This division of elements into accepted and rejected sets, in line with the theory of coherence outlined, is coherent to the extent that it reflects the maximum satisfaction of constraints possible in the network. Read, Vanman and Miller (1997) argue that such parallel constraint satisfaction processes offer a computational implementation of Gestalt principles of holistic processing.

Various equations can be used to update activation until the network settles (McClelland and Rumelhart, 1989). For example, on each cycle the activation of a unit j, a_j, can be updated according to the following equation:

$$a_j(t+1) = a_j(t)(1-d) + net_j(max - a_j(t)) \text{ if } net_j > 0, \; net_j(a_j(t) - min) \text{ otherwise}$$

Here d is a decay parameter (say .05) that decrements each unit at every cycle, *min* is a minimum activation (−1), *max* is maximum activation (1). Based on the weight w_{ij} between each unit i and j, we can calculate net_j, the net input to a unit, by:

$$net_j = \sum_i w_{ij} a_i(t).$$

Weights can be positive, representing excitatory links, or negative, representing inhibitory links.

4. Simulating Weak Coherence

Given a developed account of coherence as constraint satisfaction and an effective means of computing coherence using connectionist techniques, we can directly address the question of how weak coherence can be simulated. This section identifies a number of possibilities for impeding the integration of information in a connectionist model so that coherence fails to be maximized. These possibilities can be divided into two classes:

(1) Generation
- omit nodes
- omit links

The first class of possibilities concerns the structure of the constraint network. When a network is set up to solve a particular coherence problem, it may be that all of the required nodes and/or links between nodes are simply not generated or they are not generated in a sufficiently strong form. If this were the case, then the resulting 'gaps' in the structure could prevent the network from reaching a maximally coherent solution. However, simple omissions or weaknesses in certain parts of the knowledge net would seem to present unlikely candidates for a computational model of weak coherence, since they effectively represent knowledge deficits impacting on the performance of the system rather than the operation of an impaired coherence mechanism per se.

(2) Settling

(i) Failure to settle
- decay too low
- excitation >> inhibition

(ii) Settle too soon, producing a local maximum
- decay too high
- inhibition >> excitation
- too few cycles

A second, more interesting class of reasons why coherence might break down in a connectionist model concerns how such networks settle. Firstly, it may not settle properly due to changes in various parameters. For example, if the decay rate which serves to dampen the level of excitation present in the

network is set too low, or the levels of excitation and inhibition are tipped so that excitation is very high relative to inhibition, the inhibitory links between competing nodes may be so weak that the network is incapable of 'deciding' between them and ends up accepting incoherent alternatives. Another possibility is that the network will not settle at all, but continue to 'thrash' back and forth between the competing interpretations.

Secondly, the network may settle prematurely on a less than optimal solution. This can happen when the weights on inhibitory links are high compared to the weights on excitatory links. If the level of inhibition is too high relative to excitation, an element that is initially preferred may suppress a more coherent alternative, preventing it from becoming activated and emerging as superior. When this happens, the network may become trapped in a *local maximum*, failing to fully maximize constraint satisfaction. We suggest below that weak central coherence can be understood in terms of a neural network reaching a local rather than a global solution to a coherence problem.

Having identified these possibilities for producing weak coherence in a connectionist model, we investigated them by modeling two stimuli integration tasks that have proven difficult for autistic individuals. For each task, we constructed a constraint network and manipulated the levels of inhibition, excitation, and decay present in the network, observing the impact on task performance. The following section summarizes our finding that premature network settling produced by excessive inhibition mimics autistic thinking on the two tasks.

5. Results of Simulations

5.1 The False Belief Task

The first task modelled is a standard test for the presence of a 'theory of mind' (Wimmer and Perner, 1983), and widespread failure by autistic individuals is generally taken to indicate an absence of this social module (Baron-Cohen, Leslie and Frith, 1985). However, in addition to tapping an important component of social skills, the false belief task can also be interpreted more generally as a coherence problem that requires the simultaneous integration of multiple sources of information for its solution.

In the typical setup (Baron-Cohen et al., 1985), the child watches the unexpected transfer of a marble from a basket where the protagonist, a doll called Sally, has placed it, to a box while she is absent. The child is then asked where Sally will look for the marble on her return. The correct response recognizes that Sally will act on the basis of her false belief about the location of the marble and look in the basket, rather than where it really is, the box. Hence passing the false belief task requires the child to combine multiple pieces of information—including where the marble is, what Sally has seen, and what Sally therefore believes about the marble's whereabouts—to arrive at a correct prediction about Sally's behaviour.

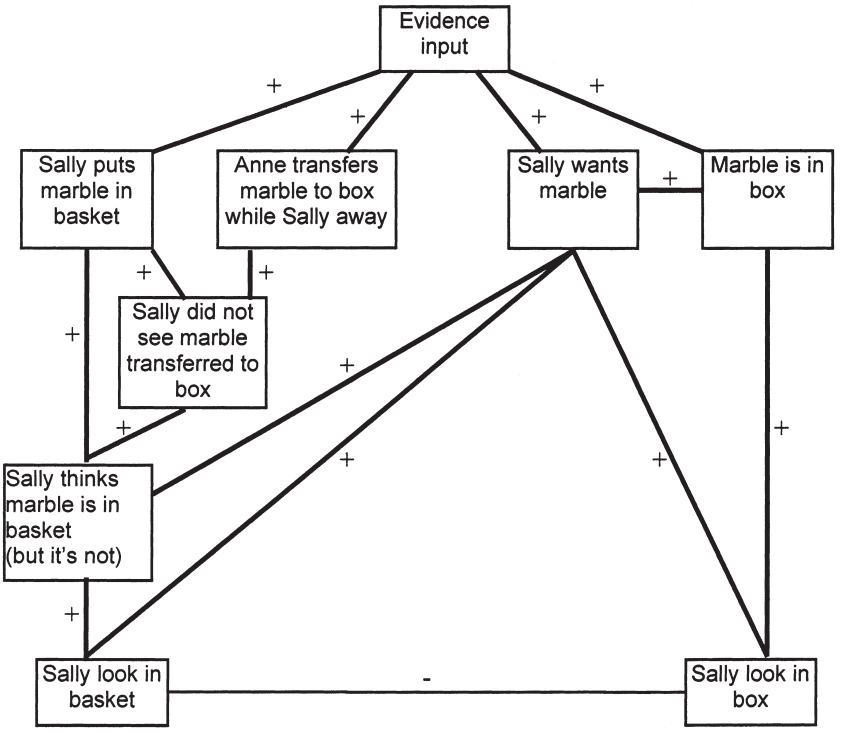

Figure 2 'Where will Sally look for her marble?' Pluses indicate excitatory links and minuses indicate inhibitory links

We interpreted the false belief task as a coherence problem that can be solved using a connectionist model displayed in Figure 2, which shows a portion of the knowledge net that would be activated when one is asked to predict Sally's behaviour. The boxes depict the nodes representing the set of elements (evidence and hypotheses) involved in the task. Bold lines indicate symmetric excitatory links between the elements, and thin lines depict symmetric inhibitory links. Coherence is computed using a programme ECHO that takes input about the explanatory and contradictory connections holding between the propositions and decides which hypotheses to accept (Thagard, 1992). The input to ECHO for the false belief task is given in the Appendix. Running ECHO on the input simulates the process by which people integrate the available information to reach the most coherent prediction about Sally's behaviour. Viewed in these terms, passing the false belief task is a matter of maximizing coherence among the set of elements so that the most coherent prediction based on Sally's attributed false belief is accepted (activated), and the competing prediction based on the marble's actual location is rejected (deactivated). Conversely, *failing* the task amounts to a breakdown in this constraint satisfaction process, producing the opposite result.

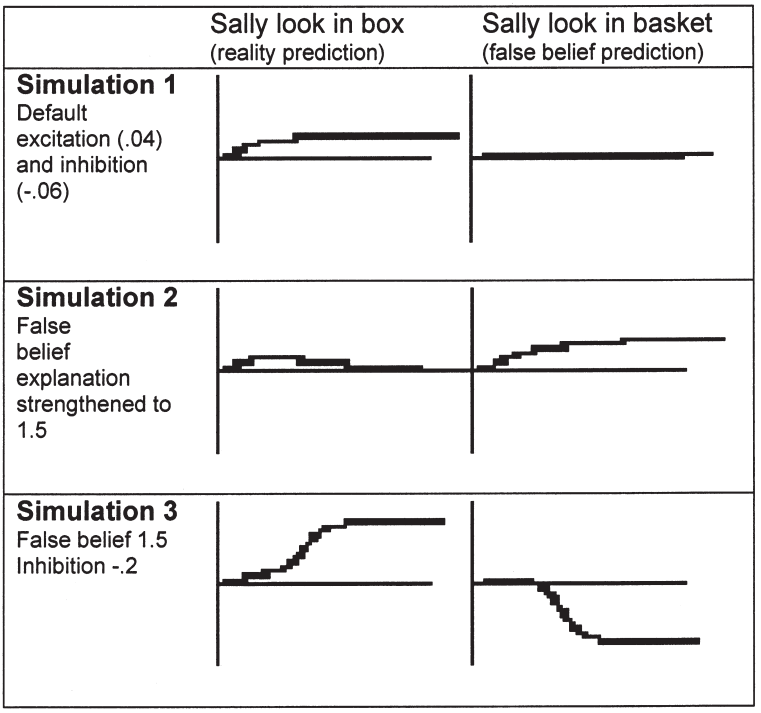

Figure 3 Results of modelling the false belief task

The results of three successive simulations of the false belief task are summarized in Figure 3. These graphs show the activation history of the two nodes representing the predictions about where Sally will look for her marble over the number of cycles it took the network to settle. The activation scale on the vertical axis ranges from −1 to 1, with the horizontal line indicating the initial activation of zero.

In Simulation 1 with parameters set at default values, *Sally look in box* is activated and the more coherent prediction *Sally look in basket* is deactivated. The network fails the task, accepting the prediction that has direct links to the evidence nodes and rejecting the prediction that is mediated by inferences concerning Sally's mental state. Behaviour of the network corresponds to the thinking of children under 4, who routinely fail the task also because of an inability to reason well about people's thought processes. The reason that ECHO prefers *Sally look in box* over *Sally look in basket* is that the former gets activation directly from the evidence nodes *Sally wants marble* and *Marble is in box*, while the activation spreads more slowly to the latter through intervening hypotheses. The more immediately obvious explanation wins out.

To counter the potency of the reality-based prediction in very young children, the strength of the explanatory relation between the nodes representing

Sally's false belief, desire, and concomitant behaviour (*look in basket*) needs to be increased. Accordingly, we increased the explanation strength from a default value of 1 to 1.5. Having incorporated an improved 'theory of mind' into the structure of the model, the outcome is reversed (Simulation 2): *Sally look in box* is deactivated and the more coherent prediction *Sally look in basket* is activated. Four-year-old children are usually capable of making this more coherent but indirect inference.

With a baseline of the network passing the false belief task established, we investigated the class of possibilities identified in section 4 for simulating weak coherence. In ECHO, the default excitation level for positive weights is .04, and the default inhibition level for negative weights is −.06. Simulation 3 shows the outcome of increasing the level of inhibition from a default value of −.06 to −.2. This manipulation significantly impedes prior task performance: *Sally look in box* is strongly activated and the more coherent prediction *Sally look in basket* is now strongly deactivated. The same result is achieved by keeping inhibition stable at −.06 but reducing excitation to .01 or below. In both cases, the coherence calculation is degraded by making excitation relatively weak compared to inhibition.

In all three simulations, the reality-based prediction gets activation directly from the evidence nodes, giving it a 'head start' in the competition for activation. In Simulation 1, this prediction remains activated because of what is effectively a knowledge deficit; that is, the initial structure of the constraint network does not represent the greater weight given to people's beliefs over reality when determining behaviour. As such, this failure falls within the first category of reasons why coherence might break down in a connectionist model identified in section 4. Accordingly, when the explanatory power of Sally's false belief is strengthened in Simulation 2, this enables coherence to be maximized among the set of elements. *Sally look in box* is deactivated over the course of the run as the more coherent prediction *Sally look in basket* steadily gains activation from its excitatory links with other nodes. However, when inhibition is subsequently increased relative to excitation (Simulation 3), the inhibitory link between the competing predictions is so strong that the prediction that is initially preferred holds down the competing (more coherent) hypothesis, preventing it from becoming activated.

Simulation 3 of the false belief task then, demonstrates one means by which weak coherence can be reproduced in a connectionist model. When inhibition is increased (disrupting the balance of inhibition/excitation in the network), this 'short-circuits' the coherence calculation and traps the network in a local maximum. As a result, only a local (here the most obvious or literal interpretation) as opposed to global solution to the coherence problem is reached. This model of weak coherence can produce failure on the false belief task and can be differentiated from failure due simply to an inadequate 'theory of mind'/network structure explanation.

In demonstrating our model of weak coherence, we have highlighted the

context dependency of false belief predictions. However, as indicated earlier, autistic failure on the false belief task together with their social difficulties more generally, are typically explained by postulating a selective impairment in a specialized 'theory of mind' mechanism (Baron-Cohen, Leslie and Frith, 1985; see also Baron-Cohen, 1995; Leslie and Roth, 1993). According to this view it is the mental content of the false belief task that presents problems for autistic individuals, rather than the requirement for integrative processing. Taken at face value then, this task would not seem the most obvious choice for attempts to model weak coherence. In support of our selection of this task, we draw attention to the following.

Firstly, by interpreting the false belief task as a coherence problem and relating autistic failure on this task to the operation of weak coherence, we do not mean to simply equate performance by autistic individuals with that of normal 3-year-olds who also typically fail the task. Rather, we demonstrated how the reasons for failure in Simulations 1 and 3 (attributed to 3-year-olds and autistic individuals respectively) are qualitatively different. Only in Simulation 3 does the operation of the network reflect an online coherence-processing problem suggestive of weak coherence.

Secondly, concerning social communication more generally, important aspects of social perception and social interaction are plausibly interpreted in terms of coherence maximization understood as a constraint satisfaction process. For example, Kunda and Thagard (1996) have demonstrated that the integration of stereotypes, traits and behaviours that occurs when we form impressions of other people, can be successfully modelled using constraint networks. Similarly, Thagard's (1992) model of explanatory coherence has been shown to have important implications for the understanding of social explanation (Read and Marcus-Newhall, 1993; Read and Miller, 1993). Collectively, such studies point to a central role for coherence mechanisms in social cognition.

Thirdly, in contrast to claims for modular deficits in autism, Frith and Happé (1994; Happé, 1999) suggest that weak central coherence is a pervasive characteristic of information processing best described as a cognitive style: 'that a different, rather than merely deficient, mind lies at the centre of autism' (Happé, 1999, p. 217). Given that this characterization implies far-reaching consequences for the processing of information, and given the role of coherence in social communication outlined above, it would seem odd if a weakened capacity for coherence was *not* affecting autistic performance on social reasoning task at some level. Therefore, rather than discounting a role for weak coherence in social cognition, the question may well be how a specific impairment in 'theory of mind' and a more general disturbance in coherence computation could possibly interact (Happé, 1999). Our simulations encourage debate on this issue by highlighting the context-dependent nature of false belief predictions over and above the specific mental content of the task, and by demonstrating how, in the presence of weak coherence, performance on the

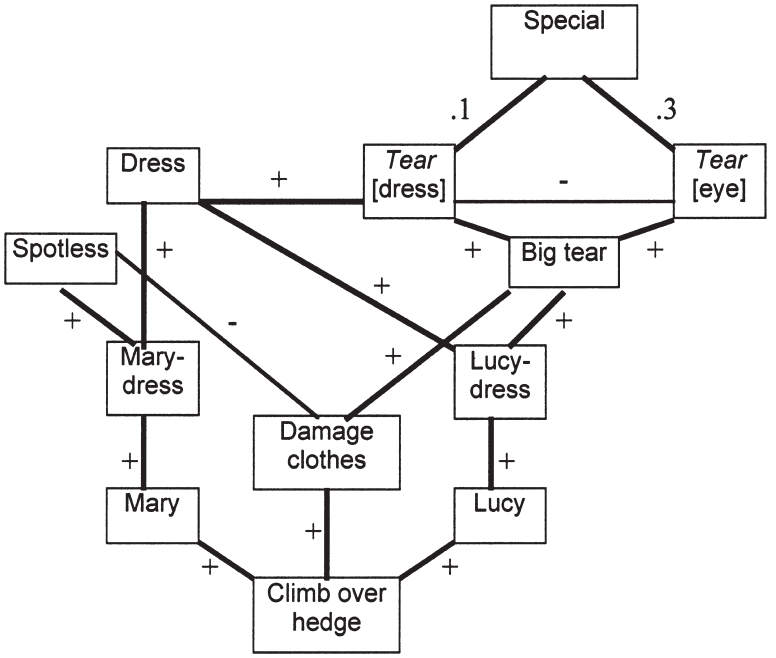

Figure 4 Disambiguating the homograph Tear

false belief task is impaired. Can this same model also mirror the cognitive problems of autistic subjects on other tasks that have been taken as a direct measure of weak central coherence?

5.2 The Homograph Task

One such task tests the ability to pronounce selected homographs (e.g. *tear* in eye vs. *tear* in dress), while reading aloud a series of sentences (Frith and Snowling, 1983; Happé, 1997; Snowling and Frith, 1986; see also Cohen and Servan-Schreiber, 1992, for an alternative model of a similar lexical disambiguation task). In order to disambiguate each homograph, subjects must make use of information supplied by the sentence in which the target word is embedded. Findings that autistic subjects fail to use sentence context to inform pronunciation of homographs (Happé, 1997), and frequently respond with the more common pronunciation of the word irrespective of context (Frith and Snowling, 1983), are cited in support of the weak coherence hypothesis as they demonstrate an inability to integrate information in order to derive meaning.

Applying our model of weak coherence to the homograph task, Figure 4 depicts part of the knowledge net that would be activated when the following sample sentence containing the homograph *tear* is presented:

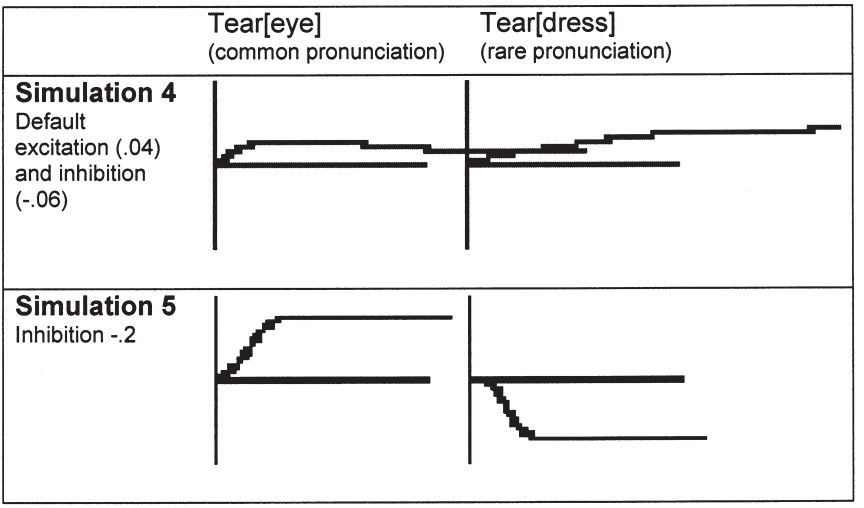

Figure 5 Results of modeling the homograph task

> The girls were climbing over the hedge. Mary's dress remained spotless, but in Lucy's dress there was a big tear. (Snowling and Frith, 1986, p. 411)

In order to solve this coherence problem and correctly disambiguate the target word, the network must integrate the relevant information provided by the sentence context so that the rare but contextually appropriate pronunciation *tear*[dress] is activated (accepted) and the more common but inconsistent pronunciation *tear*[eye] is deactivated (rejected). If however, as is suggested in the autistic case, the network's ability to compute coherence is severely compromised (by increasing the level of inhibition relative to excitation), then a maximally coherent solution is unlikely to be reached and the opposite result should obtain. We computed coherence using IMP (Kunda and Thagard, 1996), a programme that implements an associative net in which the connections between elements are expressed as positive and negative associations, rather than strict relations of explanation and contradiction. In order to run this simulation, relative frequencies for rare and common pronunciations of *tear* have been estimated (.1 and .3 respectively). The Appendix shows the input to IMP used in the homograph task.

Figure 5 summarizes our results. In Simulation 4, with coherence calculations operating normally (excitation default .04 and inhibition default –.06), the network makes use of the sentence context to correctly disambiguate the homograph: *tear*[dress] is activated more than *tear*[eye]. However, in Simulation 5, when the level of inhibition was increased to –.2, the value used in Simulation 3, the outcome is reversed: sentence context is ignored and the

more common (but less coherent) pronunciation *tear*[eye] is activated instead and *tear*[dress] is deactivated.

Therefore, as with the false belief task, increasing the level of inhibition relative to excitation impedes the network's ability to perform coherence calculations efficiently. For both tasks, this manipulation results in the more immediate or obvious judgement being accepted, rather than the judgement that makes the most sense (is maximally coherent) given the available information. As in the false belief simulation, reducing the excitation to .01 has the same effect as increasing inhibition. What matters is the relative strength of excitation and inhibition, with weak coherence performance arising when excitation is too weak compared to inhibition.

6. Conclusion

Frith's theory about the cognitive impairment in autism proposes that autistic individuals display a weakened capacity for coherence-based inference. Our simulations fill out this qualitative account by providing a computational implementation of weak coherence in a connectionist network. We found that increasing the level of inhibition relative to excitation in networks impaired their ability to maximize coherence. Strong inhibition forced the system to settle prematurely before the coherence relations between elements could take effect, so that only a local solution to the coherence problem was reached. Short-circuiting the constraint satisfaction process in this manner produces outcomes on both the false belief task and the homograph task that correspond to the performance of individuals with autism.

Our simulations show how the integration of information can break down and how fragmented judgements can occur as the result of impaired coherence calculations. While this model is clearly concerned with dysfunction at the cognitive level, such a demonstration of the breakdown of coherence in a connectionist network affords some preliminary speculation about the brain systems underpinning weak coherence. Specifically, what we have modelled is a collapse of the balance holding between excitation and inhibition in the network that normally facilitates integrative processing. By artificially creating an environment where inhibition and excitation are no longer holding each other in check, the ability of the network to maximize coherence among its elements is severely compromised. Given the destabilizing influence of excessive inhibition on our network's ability to integrate information in a meaningful fashion, it is interesting to note that a recent neural circuit theory of autism suggests excessive inhibitory lateral feedback synaptic connection strengths can impair the development of feature maps (Gustafsson, 1997). Inadequate feature maps are in turn thought to impede the coherent processing of information and hence are seen to offer a possible neural level explanation for Frith's theory of weak central coherence.

While speculations regarding neurological correlates of excessive inhibition

remain tentative, our model does provide a clear basis for new behavioural experiments. To take one example, Gentner and Toupin (1986) found that analogical mapping can fail in young children who are swamped by surface similarity information and do not see the coherent structure of analogs. We would expect autistic individuals to also fail this task, for reasons due primarily to a weakened capacity for coherence-based inference. Other coherence tasks, such as the many phenomena involving impression formation modelled by Kunda and Thagard (1996), should also display impairments deriving from weak coherence. Thus Frith's weak central coherence theory and the theory of coherence as constraint satisfaction conjoin naturally to provide a framework for future investigations of autism.

University of Canterbury
Christchurch, New Zealand

University of Waterloo
Ontario, Canada

Appendix

Input to ECHO for false belief task

```
; Evidence:
(proposition 'E1 'Sally puts marble in basket.')
(proposition 'E2 'Anne transfers marble to box while Sally away.')
(proposition 'E3 'Sally wants marble.')
(proposition 'E4 'Marble is in box.')

; False belief hypotheses:
(proposition 'FH1 'Sally did not see marble transferred to box.')
(proposition 'FH2 'Sally thinks marble is in basket (but it's not).')
(proposition 'FH3 'Sally look in basket.')

; Reality-based hypotheses:
(proposition 'RH1 'Sally look in box.')

; Contradictions:
(contradict 'FH3 'RH1)

; False belief explanations:
(explain '(E2) 'FH1)
(explain '(E1 FH1) 'FH2)
(explain '(FH2 E3) 'FH3) [strengthened to 1.5 for simulations 2 and 3]
```

```
; Reality-based explanations:
(explain '(E3 E4) 'RH1)

(data '(E1 E2 E3 E4))
```

Input to IMP for homograph task

```
(observed 'Tear 'Tear-dress .1)
(observed 'Tear 'Tear-eye .3)
(observed 'Mary 'climb-over-hedge)
(observed 'Lucy 'climb-over-hedge)
(observed 'Mary-dress 'spotless)
(observed 'Lucy-dress 'big-tear)
(associate 'climb-over-hedge 'damage-clothes)
(associate 'Mary 'Mary-dress)
(associate 'spotless 'damage-clothes-1)
(associate 'Mary-dress 'dress)
(associate 'Lucy 'Lucy-dress)
(associate 'big tear 'damage-clothes)
(associate 'Lucy-dress 'dress)
(associate 'big-tear 'Tear-dress)
(associate 'big-tear 'Tear-eye)
(associate 'Tear-dress 'Tear-eye-1)
(associate 'Tear-dress 'dress)
```

References

Baron-Cohen, S. 1995: *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A.M. and Frith, U. 1985: Does the autistic child have a 'theory of mind'? *Cognition*, 21, 37–46.

Cohen, J.D. and Servan-Schreiber, D. 1992: Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, 99, 45–77.

Frith, U. 1970a: Studies in pattern detection in normal and autistic children: I. immediate recall of auditory sequences. *Journal of Abnormal Psychology*, 76, 413–20.

Frith, U. 1970b: Studies in pattern detection in normal and autistic children: II. reproduction and production of color sequences. *Journal of Experimental Child Psychology*, 10, 120–35.

Frith, U. 1989: *Autism: Explaining the Enigma*. Oxford: Basil Blackwell.

Frith, U. and Happé, F. 1994: Autism: beyond 'theory of mind'. *Cognition*, 50, 115–32.

Frith, U. and Snowling, M. 1983: Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology*, 1, 329–42.

Gentner, D. and Toupin, C. 1986: Systematicity and surface similarity in the development of analogy. *Cognitive Science*, 10, 277–300.

Gustafsson, L. 1997: Inadequate cortical feature maps: a neural circuit theory of autism. *Biological Psychiatry*, 42, 1138–47.

Happé, F.G.E. 1996: Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. A research note. *Journal of Child Psychology and Psychiatry*, 37, 873–77.

Happé, F.G.E. 1997: Central coherence and theory of mind in autism: reading homographs in context. *British Journal of Developmental Psychology*, 15, 1–12.

Happé, F.G.E. 1999: Autism: cognitive deficit or cognitive style? *Trends in Cognitive Sciences*, 3, 216–22.

Hermelin, B. and O'Connor, N. 1970: *Psychological Experiments with Autistic Children*. Oxford: Pergamon Press.

Holyoak, K.J. and Thagard, P. 1995: *Mental Leaps: Analogy in Creative Thought*. Cambridge, MA: MIT Press/Bradford Books.

Kunda, Z. and Thagard, P. 1996: Forming impressions from stereotypes, traits, and behaviors: a parallel-constraint-satisfaction theory. *Psychological Review*, 103, 284–308.

Leslie, A.M. and Roth, D. 1993: What autism teaches us about metarepresentation. In S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen (eds), *Understanding Other Minds: Perspectives From Autism*. Oxford University Press.

McClelland, J.L. and Rumelhart, D.E. 1989: *Explorations in Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Read, S.J. and Marcus-Newhall, A.R. 1993: Explanatory coherence in social explanations: a parallel distributed processing account. *Journal of Personality and Social Psychology*, 65, 429–47.

Read, S.J. and Miller, L.C. 1993: Rapist or 'regular guy': explanatory coherence in the construction of mental models of others. *Personality and Social Psychology Bulletin*, 19, 526–40.

Read, S.J., Vanman, E.J. and Miller, L.C. 1997: Connectionism, parallel constraint satisfaction processes, and Gestalt principles: (re)introducing cognitive dynamics to social psychology. *Personality and Social Psychology Review*, 1, 26–53.

Shah, A. and Frith, U. 1983: An islet of ability in autistic children: a research note. *Journal of Child Psychology and Psychiatry*, 24, 613–20.

Shah, A. and Frith, U. 1993: Why do autistic individuals show superior performance on the block design task? *Journal of Child Psychology and Psychiatry*, 34, 1351–64.

Snowling, M. and Frith, U. 1986: Comprehension in 'hyperlexic' readers. *Journal of Experimental Child Psychology*, 42, 392–415.

Thagard, P. 1992: *Conceptual Revolutions*. Princeton: Princeton University Press.

Thagard, P. forthcoming. *Coherence in Thought and Language*. Cambridge, MA: MIT Press.

Thagard, P. and Millgram, E. 1995: Inference to the best plan: a coherence theory of decision. In A. Ram and D.B. Leake (eds), *Goal-Driven Learning*. Cambridge, MA: MIT Press.

Thagard, P. and Verbeurgt, K. 1998: Coherence as constraint satisfaction. *Cognitive Science*, 22, 1–24.

Wimmer, H. and Perner, J. 1983: Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–28.