LANGUAGE AND COGNITIVE PROCESSES
0000, 00 (00), 1–25

Psychology Press
Taylor & Francis Group

The development of similarity: testing the prediction of a computational model of metaphor comprehension

Harry R. M. Purser and Michael S. C. Thomas
Developmental Neurocognition Laboratory, Birkbeck College, London, UK

Sarah Snoxall and Denis Mareschal
Centre for Brain and Child Development, School of Psychology, Birkbeck College, London, UK

An empirical study is presented that tests a novel prediction generated by the Metaphor-by-Pattern-Completion (MPC) connectionist model of metaphor comprehension (Thomas & Mareschal, 2001). The MPC model predicts a developmental progression in the way that children process metaphors, from a preference for basic-level metaphors to a preference for subordinate-level metaphors. Preference for different kinds of verbal similarity statements was assessed for 73 children, aged 4–10, along with justifications. The prediction of the model was confirmed, providing evidence for the attendant assumption of the model, specifically that metaphorical comprehension is intimately linked to the emerging structure of semantic representations in children.

Keywords: Connectionism; Development; Metaphor; Modelling.

INTRODUCTION

Although developmental literature has documented many examples of younger children's figurative language (e.g., Billow, 1981; Winner, 1979), it has been claimed that children cannot understand metaphorical speech until they are relatively old (Inhelder & Piaget, 1964; Piaget, 1962). A possible reason for this view is that metaphorical proficiency relies on

---

Correspondence should be addressed to Dr Harry Purser, Developmental Neurocognition Laboratory, School of Psychology, Birkbeck College, Malet Street, London WC1E 7HX, UK. E-mail: h.purser@bbk.ac.uk

This research was supported by European Commission grant NEST-029088 (ANALOGY), British Academy grant SG30602, and MRC grant G0300188.

http://www.psypress.com/lcp DOI: 10.1080/01690960902786530

many component abilities, such as metalinguistic skill, semantic knowledge, and a capacity for communicative pragmatics (see Vosniadou, 1987). Setting aside the hypothesis that utterances appearing to be metaphors might, in fact, be linguistic errors such as overextensions of word meanings (e.g., Chukovsky, 1968), which can be ruled out by checking that the child knows the actual name of the object referred to (see Gardner, Winner, Bechhofer, & Wolf, 1978), one should be clear on how apparent metaphor in young children is be interpreted. For an utterance to be deemed metaphorical, the child must point towards some similarity between objects in different categories, and at the same time be aware on some level that the objects belong to different conceptual categories.

The notion that figurative language ability is constrained by conceptual knowledge is supported by evidence that children tend to produce metaphors in domains with which they are more familiar (Gottfried, 1997). Research in the closely related field of analogical reasoning supports this claim (Goswami, 1996), suggesting that limitations in reasoning are intimately tied to having the requisite knowledge to demonstrate this ability. Keil (1986) found that if children understood a metaphor from a particular conceptual domain, they tended to understand all the other metaphors tested from that domain, but metaphors from other domains would come to be understood at different times. Furthermore, there appeared to be a developmental progression in the kinds of metaphor that were understood, with those referring to animate/inanimate distinctions, such as 'the car is hungry', understood before those based on physical/non-physical distinctions, such as 'the idea is ripe'. Keil (1986) concluded that metaphor comprehension does not emerge as a specific linguistic or processing ability at a certain age, but instead develops on a domain-by-domain basis, reflecting the amount and quality of conceptual knowledge in the domains juxtaposed by the metaphor. Thus, the development of metaphorical cognition is placed within the context of conceptual development.

Using such a framework, Vosniadou and Ortony (1983) investigated the preference of 3–6-year-old children, and adults, for three types of similarity statement: literal ('a river is a lake'), metaphorical ('a river is a snake'), and anomalous ('a river is a cat'). The authors found that even the youngest children were able to distinguish literal and metaphorical statements from anomalous ones, suggesting that children as young as 3 years old consider both literal and metaphorical statements to be meaningful similarity statements, as distinct from anomalous statements. However, only participants aged 4 years old and upwards were able to distinguish further between literal and metaphorical comparisons.

In addition to this comparison task, a different group of participants undertook a categorisation task that tested knowledge of the categories to which the terms of the comparison task belong. It was found that only children aged 4 years and upwards demonstrated appropriate category knowledge. The authors concluded that while 3-year-olds process metaphors as meaningful similarity statements, they are not aware that the terms belong to different conventional categories. Thus, the study suggested that children do not understand metaphorical similarity until 4 years of age. Marschark and Nall (1985) have further related these findings to gradual formation of conceptual categories and suggested that because a young child's conceptual knowledge may not yet have clustered into clearly defined categories, a statement that is considered metaphorical by adults may be taken literally by the child. The instances in which young children appear to produce or comprehend metaphors may, in fact, reflect overlapping or poorly delineated conceptual categories.

Glucksberg and Keysar (1990) have postulated the view that metaphors are, in fact, implicit categorisation, or class-inclusion, statements, in which the topic of a metaphor is reclassified as the same kind of thing as the vehicle. For example, in 'my job is a jail', 'job' is assigned to the category of restraining, confining and unpleasant situations, of which 'jail' is a prototypical member. When this ad hoc categorisation process occurs within a relevant context, it results in a meaningful comparison and the topic is re-evaluated in light of the features associated with the vehicle category. In this way, attributes of members of the category typified by 'jail' are transferred to 'job'. More recently, Bowdle and Gentner (2005) have suggested that such metaphorical categories may be created during the development of metaphor comprehension, emerging as abstract relational schemas derived from common relational structures of target and vehicle concepts.

Rosch and colleagues (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) have suggested that categories develop hierarchically and, over time, form three distinct levels: a *basic* level, across which the differences between concepts are maximised, a more specific or *subordinate* level, and a more general or *superordinate* level. For example, where 'animal' represents the superordinate level, a basic-level term might be 'dog', with 'Labrador' at the subordinate level (see Rogers & Patterson, 2007, for a discussion of basic-level effects). More recent research has indicated that infants learn subordinate-level categories after basic-level ones, facilitating response to increasingly subtle and complex patterns of similarity (see Mandler, 2007, for an overview).

Thomas and Mareschal (2001) proposed a computational implementation called the Metaphor by Pattern Completion model (MPC), which captures some aspects of basic- and subordinate-level representation, and, importantly, specified the mechanisms of change underlying the development of similarity comparisons. The essence of MPC is that similarity is viewed as a transformational process, rather than a process of counting features that are common or different between two representations (see Thomas & Mareschal, 1997; see also Hahn, Chater, & Richardson, 2003). At the core of the MPC is a connectionist model of semantic memory. Thomas and Mareschal proposed that comparison processes constitute a strategic use of semantic memory enabled by language. As a result, the model directly links the development of metaphor comprehension to the development of semantic representations.

There are several computational models of metaphor comprehension (and analogical reasoning) that predate MPC, but all pertain to the adult state. Although connectionist networks are highly suited to simulating developmental phenomena, none of the previous models that used this architecture involved a developmental aspect: earlier models of metaphor comprehension only focused on the soft constraint satisfaction abilities of connectionist networks to simulate the interactions of semantic domains when they are juxtaposed in comparisons. Some modellers have focused on structural mapping accounts of analogy formation, such as structure-mapping theory (SMT; Gentner, 1983, 1989), in which target and vehicle domains are compared by aligning aspects of their relational structure and evaluating shared attributes. For example, the Analogical Constraint Mapping Engine (ACME; Holyoak & Thagard, 1989) and Learning and Inference with Schemas and Analogies (LISA; Hummel & Holyoak, 1997) focus mainly on the process of mapping between domains, with multiple constraint satisfaction of both structural and semantic information. Both are hybrid connectionist/symbolic systems. ACME uses structural, semantic, and pragmatic constraints together, in order to make decisions about local correspondences between elements of relational structure that aim to produce an overall mapping which is psychologically plausible. LISA may be considered an extension of ACME in which targets and vehicles are bound together by temporal synchrony.

Purely connectionist models have focused on the ability of microfeature representations of concepts to simulate interactions between knowledge bases (e.g., Chandler, 1991; Sun, 1995; Thomas & Mareschal, 2001). For example, Sun's (1995) CONSYDERR model is a three-layer feed-forward backpropagation network, with the input layer supporting localist representations of concepts as individual nodes and the middle layer representing concepts as microfeatures. Despite the success of such models in capturing some aspects of metaphor comprehension, they have tended to rely on domain-specific representations that are largely pre-structured, which prevents such models from being used to investigate how representations might emerge developmentally. More recently, Leech, Mareschal, and Cooper (2008) have extended the assumptions of the MPC model to consider relational mapping in analogical reasoning, also within a developmental context. Here again, the key assumptions are that similarity comparisons are transformations, and the core of the model is a semantic memory system not specifically designed (or trained) to perform analogies.

Thomas and Mareschal's (2001) developmental rendering of MPC focused only on attribute mapping in metaphor comprehension, setting aside issues of structural alignment. Although a weakness of this approach is that it limits the scope of the metaphors to which the model can be applied, its strength is that it allows initial steps to be taken towards exploring the developmental aspects of metaphor comprehension, in the context of semantic development. Despite the limitations imposed by simplification, implemented models offer several key advantages. These include the clarification of existing verbal theories through the greater level of detail required in implementation, the demonstration of the sufficiency of a given set of theoretical assumptions to generate qualitative and potentially quantitative patterns of behaviour, and, perhaps most importantly, the opportunity to derive and test novel predictions about behaviour.

In the remainder of this paper, we present a modelling section, which details an implementation of MPC, an empirical section, in which a developmental prediction of MPC is tested, and a general discussion section.

COMPUTATIONAL MODELLING OF METAPHOR COMPREHENSION

MPC model

Full details of the MPC model, along with an evaluation of its key assumptions, are given in Thomas and Mareschal (2001). In essence, the MPC model suggests that on hearing a simple metaphor such as 'The apple is a ball', the listener attempts to categorise 'apple' as a kind of 'ball' and in so doing alters the representation of 'apple' such that its features become more consistent with those of balls. This process is conceived as having two distinct stages. In line with Glucksberg and Keysar's (1990) view of metaphor comprehension as a categorisation process, the first stage consists of misclassification of a semantic input. The metaphor above, where 'apple' is the topic and 'ball' is the vehicle, is understood by inputting a representation of 'apple' to an autoassociator network trained on exemplars of balls. Categorisation is evaluated on the basis of the accuracy with which 'apple' is reproduced by the network (see McClelland & Rumelhart, 1986, on the use of autoassociator networks as a model of semantic memory).

The output of the network will be a representation of 'apple' transformed to make it more consistent with the knowledge the network has stored about balls. This transformation reflects covariant structure of features of 'apple' with particular features of balls, such that 'apple' will inherit further features by a process of pattern completion. This inheritance of features depends on both terms of the metaphor, providing an implementation of Black's (1979) interaction theory of metaphor comprehension. In the second stage of comprehension, the degree to which the meaning of the input has been changed is compared to the *expected* amount of change given the current discourse context (Vosniadou, 1989). Depending on this context, the statement inputted to the network is then classified as literal, metaphorical, or anomalous. For this example, the discourse context determines whether the statement 'The apple is a ball' is viewed as a literal correction to the misidentification of an object (i.e., where the discourse context is such that a high meaning change is expected), a metaphor highlighting the roundness and throwability of a given apple (i.e., where only the enhancement of certain features is accepted, reflecting the communicative intent of a metaphor; where an intermediate meaning change is expected), or an erroneous misidentification of an apple (i.e., a low expected meaning change is expected).

In order to assess whether A is a member of B, A is transformed by B knowledge. If it is little changed, it is likely a member of B. Reproduction as a means of assessing category membership is a widely used mechanism in connectionist models of memory (see Mareschal & Thomas, 2007).

The model is able to account for a number of empirical phenomena (Thomas & Mareschal, 2001), such as non-reversibility of metaphorical comparisons and the predictability of interactions between topic and vehicle. In addition, the degree and nature of metaphorical semantic transformations depends on the amount and quality of information stored by the semantic network. In this way, the model makes a link between metaphor comprehension and semantic development.

A developmental implementation of MPC

Building on this simple example, the development of metaphor comprehension by MPC was traced as training progressed in the knowledge base *types of ball*. A network with 16 input units, 16 output units, and 10 hidden units was trained to autoassociate a set of input patterns that defined the semantic knowledge of the vehicle domain (see Figure 1). The model used 10 hidden units, a number great enough to allow good training performance, but small enough to encourage generalisation. All processing units in the network had sigmoid activation functions.

The network was trained for 500 presentations of the complete training set. At each epoch, the training set was presented in a different random order. The learning rate and momentum were set to 0.05 and 0.0, respectively. Metaphor comprehension performance was evaluated at 0, 1, 2, 3, 4, 5, 7, 10, 15, 20, 30, 45, 70, 110, 200, and 500 epochs of training. The

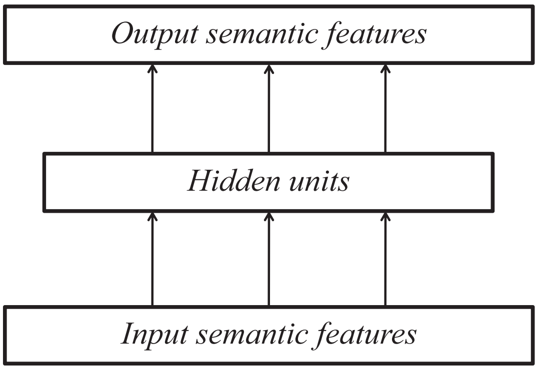

Figure 1. Architecture of the MPC model.

results reflect an average over 10 replications with different initial random seeds.

The training set was constructed around eight prototypes of various types of ball, constituting the 'ball' knowledge base. Prototypes were defined over five clusters of features: colour (red, green, brown, white), shape (round, irregular), consistency (soft, hard), size (small, large), weight (heavy, light), and associated action (thrown, kicked, hit, eaten), for a total of 16 semantic features. The last feature was included to permit anomalous and metaphorical comparisons. It was assumed that, in general, all concepts can be described by subsets of the same large feature set, and that the organisation of knowledge into different categories occurs within the hidden unit representations through learning. Feature values ranged between 0 and 1, so that the higher the activation, the more prominent the feature. Opposite feature values (e.g., small and large) were encoded on separate inputs to allow the coding of an absence of knowledge. From each prototype, 10 exemplars were generated by adding Gaussian noise (with standard deviation of 0.35) to the prototype pattern. The final training set thus constituted 80 exemplars of balls (8 prototypes $\times$ 10 exemplars), which corresponded to the individual's experience with the domain of balls.

Assessing different semantic comparisons

Comparisons were evaluated by inputting novel exemplars to the network and examining their reproduction at output. The more accurate each reproduction, the greater the similarity of the novel item to the knowledge stored within the network. Nine test comparisons were performed, which fell into three classes: (1) literal comparisons, (2) metaphorical comparisons, and (3) anomalous comparisons. For each comparison, novel exemplars were used. Literal comparisons involved exemplars of balls close to the prototypical values. Metaphorical comparisons involved inputs that shared some features with balls, but differed on other features. Anomalous comparisons involved exemplars that shared few features with balls.

The status of the input vectors for the different classes of comparison was established by comparing the novel input with the ball prototype vectors used to generate the training set (see Table 1, upper section). This was achieved by computing the angle between the two vectors in semantic space (see Thomas & Mareschal, 1997). For the literal comparisons, angles were chosen to be less than 10 degrees; for the metaphorical comparisons, they were between 40 and 45 degrees; and for the anomalous comparisons, between 60 and 66 degrees. (An angle of 90 degrees would constitute a novel pattern orthogonal to, or completely different from, all the prototypes used to generate the exemplars in the knowledge base.) Novel comparisons are shown in Table 1, lower section. A perfect reproduction of the input at the output would indicate a similarity of 1.0. The transformational similarity (*S*) of each novel comparison to the ball knowledge base was defined as:

$$S = 1 - RMS\ Error$$

For a vector of with elements X_1 to X_N,

$$RMS = Sqrt((X_1)^2 + (X_2)^2 + (X_3)^2 + \ldots\ldots + (X_N)^2/N)$$

RMS error in this case is the distance in Euclidean space between the input and output vectors. An *RMS* error of 0 would give a similarity of $S = 1$. High similarity implies low semantic distortion (as expected in a literal comparison), moderate similarity implies moderate semantic distortion (as expected in a metaphorical comparison), and low similarity implies high semantic distortion (as expected in an anomalous comparison). The similarity of novel comparisons was evaluated at various points during training.

Results

Figure 2 shows transformational similarity (*S*) for the three types of comparison statement (literal, metaphorical, and anomalous), as learning progressed. Although there is initially very little difference between *S* for each comparison type, after only a few learning trials (where one trial is exposure to the full set of training knowledge) the anomalous comparisons clearly separate from the literal and metaphorical ones. *S* of literal and metaphorical comparisons continues to be similar until around seven learning trials, at which point these two comparison types diverge. For the rest of training, the three comparison types remain clearly delineated.

By attempting to classify novel inputs, the network transforms them. This transformation of the representation, which alters semantic attributes, corresponds to the comprehension of a comparison statement. For metaphorical comparisons, this change of semantic attributes corresponds to an enhancement of meaning. There appear to be three distinct stages in how

TABLE 1
Upper section: Semantic feature codes for prototypes in the knowledge base. Adding noise to the prototypes creates training sets. Lower section: Novel patterns used in literal, metaphorical, and anomalous comparisons.

	Colour				*Action*				*Shape*		*Consistency*		*Size*		*Weight*	
Prototypes	*Red*	*Green*	*Brown*	*White*	*Eaten*	*Thrown*	*Hit*	*Kicked*	*Round*	*Irregular*	*Soft*	*Hard*	*Small*	*Large*	*Heavy*	*Light*
Football (white)	0.00	0.00	0.00	0.90	0.00	0.20	0.00	0.95	0.90	0.00	0.00	0.80	0.00	0.90	0.90	0.00
Football (brown)	0.00	0.00	0.90	0.00	0.00	0.20	0.00	0.95	0.90	0.00	0.00	0.80	0.00	0.90	0.90	0.00
Cricket ball	0.90	0.00	0.00	0.00	0.00	0.98	0.97	0.00	0.90	0.00	0.00	0.98	0.80	0.00	0.80	0.00
Ping-pong ball	0.00	0.00	0.00	0.95	0.00	0.10	0.98	0.00	0.98	0.00	0.00	0.98	0.95	0.00	0.00	0.95
Tennis ball	0.00	0.90	0.00	0.00	0.00	0.80	0.98	0.00	0.90	0.00	0.80	0.00	0.80	0.00	0.00	0.85
Squash ball (red)	0.80	0.00	0.00	0.00	0.00	0.50	0.98	0.00	0.93	0.00	0.85	0.00	0.95	0.00	0.00	0.90
Squash ball (green)	0.00	0.90	0.00	0.00	0.00	0.50	0.98	0.00	0.93	0.00	0.85	0.00	0.95	0.00	0.00	0.90
Beach ball	0.98	0.00	0.00	0.00	0.00	0.90	0.90	0.90	0.90	0.00	0.98	0.00	0.00	0.98	0.00	0.90
Novel comparisons																
Literal:																
Football (white)	0.00	0.00	0.00	0.85	0.00	0.20	0.00	0.98	0.80	0.00	0.10	0.80	0.00	0.90	0.80	0.00
Beach ball	0.90	0.00	0.00	0.00	0.00	0.80	0.70	0.90	0.70	0.00	0.90	0.00	0.00	1.00	0.00	0.80
Ping-Pong ball	0.00	0.00	0.00	0.99	0.00	0.20	0.99	0.00	0.95	0.00	0.00	0.90	0.95	0.00	0.00	0.97
Metaphorical:																
Apple (red)	0.80	0.00	0.00	0.00	0.95	0.05	0.00	0.00	0.75	0.15	0.70	0.20	0.70	0.00	0.00	0.50
Pumpkin	0.20	0.00	0.70	0.00	0.80	0.00	0.00	0.00	0.80	0.50	0.80	0.60	0.00	0.80	0.90	0.00
Apple (green)	0.00	0.95	0.00	0.00	0.95	0.05	0.00	0.00	0.75	0.15	0.70	0.20	0.70	0.00	0.00	0.50
Anomalous:																
Kite	0.99	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.99	0.00	0.98	0.00	0.95	0.20	0.80
Spaghetti	0.00	0.00	0.80	0.20	0.97	0.00	0.00	0.00	0.00	0.70	0.80	0.20	0.00	0.70	0.00	0.60
Toast	0.00	0.00	0.80	0.10	0.80	0.00	0.00	0.00	0.00	0.80	0.80	0.00	0.80	0.00	0.00	0.90

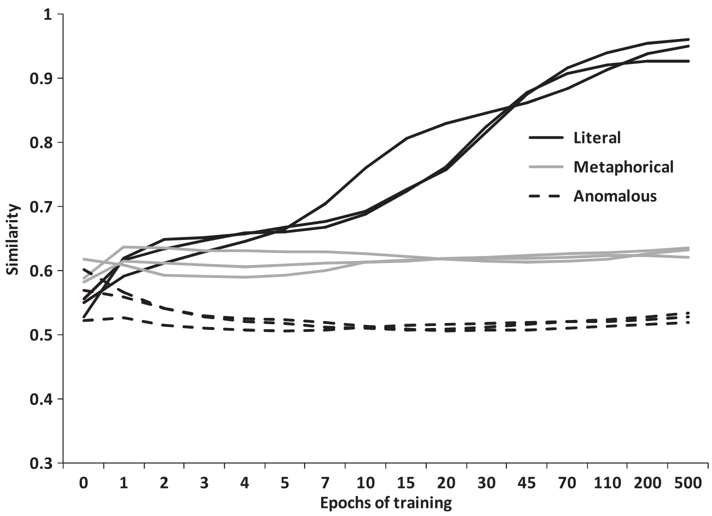

Figure 2. Transformational similarity (*S*) of test comparisons to the ball knowledge base stored by the network. Three examples each are given for literal, metaphorical, and anomalous comparisons. A discourse context of intermediate expected meaning change is assumed.

MPC responds to metaphorical comparisons across training. In the first, there is poor pattern completion, owing to a relatively impoverished knowledge base. Second, there is some enhancement of semantic features of the topics, as the amount and quality of information stored by the network increases. In the examples of ‘The apple is a ball’ and ‘The pumpkin is a ball’, there is transfer of prototypical features of balls to these fruit topics. This enhancement occurs according to an early, prototypical notion of ball, drawing on an average of all exemplars of balls and corresponding to the *basic* level of that category. For example, on average, most balls (in the set of exemplars used here) are hit, rather than thrown or kicked, so all firm and round objects inherit this feature. In the third phase, however, as further training results in more distinct representations of balls within the knowledge base, features are transferred from the type of ball that is most similar to the object inputted to the network, corresponding to the *subordinate* level of the ball category.

Figure 3a and 3b show transformed feature values for inputs ‘apple’ and ‘pumpkin’, respectively. At 7 epochs of training, ‘apple’ and ‘pumpkin’ have similar values for the features ‘hit’, ‘thrown’, and ‘kicked’, loading higher on ‘hit’ than on the others. Both topics are transformed by a poorly delineated notion of ball. However, by 100 epochs, the features that are enhanced by the network are determined by the size of the input objects: while the feature loadings of ‘apple’ are much the same as earlier in training, with ‘hit’ highly activated, those for ‘pumpkin’ have shifted markedly, with ‘kicked’ now a prominent feature and ‘hit’ loaded only minimally. This is because apple is most similar to small balls that are typically hit, such as cricket or baseballs,

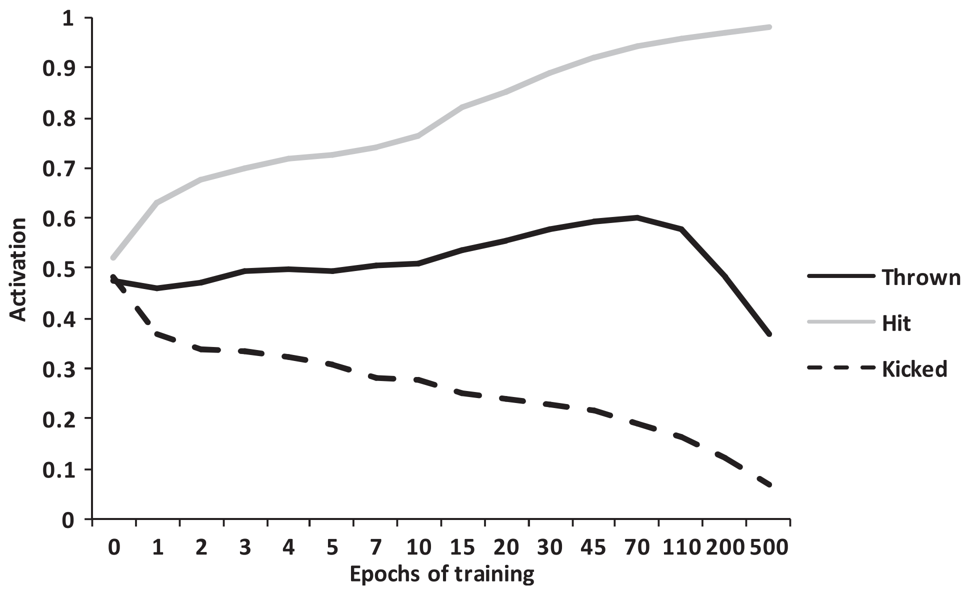

Figure 3a. Transformed output feature activations for network input 'apple' during training.

while pumpkin is most similar to larger balls that are typically kicked, such as football. In this way, MPC generated a testable prediction: *attribute inheritance will shift from basic-level to subordinate-level across development*. To our knowledge, this prediction has not been (explicitly) generated by any

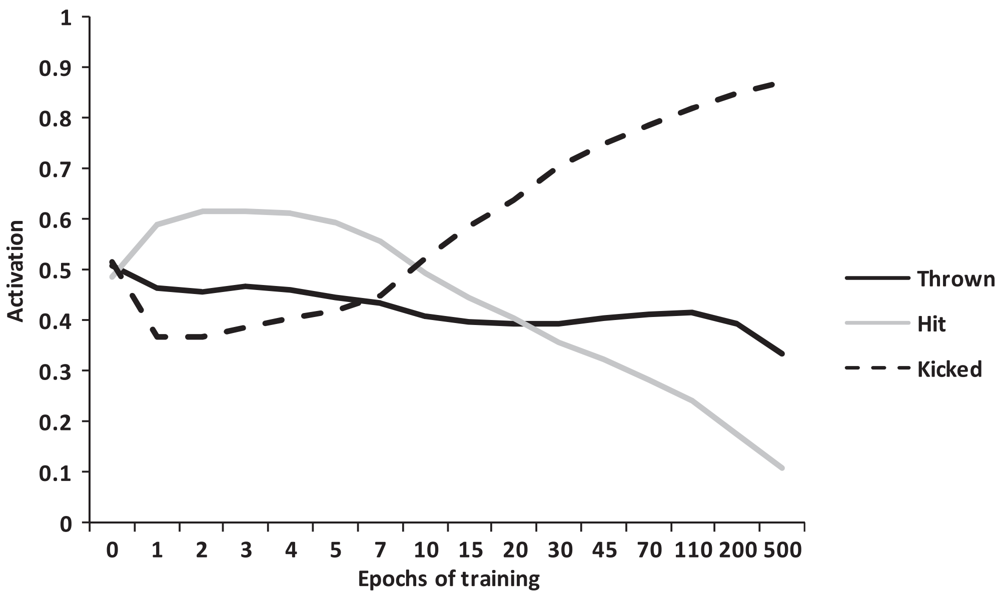

Figure 3b. Transformed output feature activations for network input 'pumpkin' during training.

previous theory or model. It should be noted that this prediction made by MPC was not a previously anticipated consequence of running the model. It is a key strength of computational models that they are capable of generating empirical predictions and help to clarify the relationships between theoretical entities (see Mareschal & Thomas, 2007).

The following section details a test of this novel prediction with children aged 4–10 years old, by investigating whether such a cognitive shift occurs in metaphor comprehension. A test of the prediction would provide evidence for or against the assumptions of the MPC model, such as that metaphorical comprehension is linked to the emerging structure of semantic representations. The prediction was evaluated in two ways. First, there was a measure of children's preference for basic- and subordinate-level metaphors, with the experimental hypothesis that younger children would prefer basic-level metaphors while older children would prefer subordinate-level ones. For isolated comparisons of this nature, it was assumed that a good metaphor is constituted by a comparison that highlights those features of the topic that are salient and diagnostic given the discourse context. Discourse contexts were created to produce the expectation of an intermediate level of meaning change. Second, children's justifications of their preferences were recorded, in order to gauge the level at which attribute inheritance occurs. That is, to demonstrate that should an increase in subordinate comparison preference occur, it is for the reasons predicted by the model. The experimental hypothesis from MPC was that younger children's justifications would involve more transfer of basic-level attributes (e.g., 'round') from topic to vehicle while older children would transfer more subordinate-level attributes (e.g., 'can throw').

EMPIRICAL TEST OF THE MODEL'S PREDICTION

Method

Participants

Three groups of typically developing children were recruited on the basis of age: twenty-eight 4–5-year-olds, twenty 7–8-year-olds, and twenty-five 9–10-year-olds. In forming these groups, three children were excluded on the basis of failing either the vocabulary test (described below) or by demonstrating a failure to understand the task. Participant details are given in Table 2.

Materials

Main task. The materials were 25 spoken metaphors, each with two possible levels of vehicle: basic and subordinate. For example, 'moon' could

TABLE 2
Participant characteristics; CA = chronological age in years and months.

	Mean CA	*Range*
4–5-year-olds	4;10	4;2–5;1
7–8-year-olds	7;7	7;3–8;10
9–10-year-olds	9;8	9;2–10;1

be paired with either 'coin' (basic) or '10p piece' (subordinate). Well-known metaphorical comparisons were avoided, in order to focus on novel metaphors. Of the 25 metaphors, 20 formed the main set of materials; the remainder formed a reserve set that could be drawn from whenever there were vocabulary difficulties with a metaphor from the main set. The metaphorical topics and vehicles used are listed in Table 3.

Participants were instructed that a stuffed toy, representing an alien character named 'Zork', was learning to speak English and had made a number of comments that were to be rated. A scoreboard with five rows of teddies was presented. Participants were instructed in the following scoring system, which was presented either from 1–5 or 5–1, randomly, by the experimenter:

- 5 teddies – very good
- 4 teddies – quite good
- 3 teddies – OK
- 2 teddies – quite bad
- 1 teddy – very bad

Following a pilot experiment involving three children from each age group, a board game was designed in order to help the younger age groups maintain concentration by marking progress through the trials. Piloting indicated that this was not necessary for the 9–10-year-old children. There were 20 moves on the board, one for every two trials. Every five moves, there was a special 'Zork-box', at which point the participant received a sticker.

Vocabulary task. Comprehension of potentially difficult words used in the main task was examined, in order to control for linguistic and conceptual knowledge. The words selected for the vocabulary test were based on age of acquisition norms from studies by Morrison, Chappell, and Ellis (1997) and Morrison and Ellis (1995). The selection of words for each age group was based on the age at which 75% of the participants (of the norming studies) correctly identified each word. Where there were no norming data for a word used in the main task, that word was used in the vocabulary test.

TABLE 3
Basic and subordinate level metaphors.

List of stimuli		
Topic	*Basic level vehicle*	*Subordinate level vehicle*
Main set		
Apple	Ball	Tennis ball
Friend in pool	Fish	Dolphin
Stars	Jewels	Diamonds
Wet hair	Pasta	Spaghetti
Sound of birds	Musical instruments	Flutes
Thin old woman	Insect	Stick insect
Scary dog barking	Dinosaur	Tyrannosaurus rex
Boy on swings	Bird	Eagle
Friend swimming	Fish	Starfish
Fingers	Sticks	Twigs
Moon	Coin	10p piece
People seen from a distance	Insects	Ants
Sound of a motorbike	Bug	Bee
Bristles on broom	Hair	Fringe
Blanket	Bear	Teddy bear
Beads on necklace	Sweets	Smarties
Cat purring	Car	Racing car
Big tall man	Building	Tower
Tyres on car	Shoes	Trainers
Sponge	Cloud	Raincloud
Reserve set		
Toddler playing	Cat	Kitten
Chimney	Hat	Top hat
Strong dad	Ape	Gorilla
Sun	Fruit	Orange
Gentle girl	Sheep	Lamb

On each trial, a series of four pictures at a time was presented, arranged in a row. Each row consisted of a picture of the target word, in the context of three distractors. The participant was simply required to select the picture from each series of four that corresponded with the spoken target word given by the experimenter. Basic level words were tested in the context of words at the same level (e.g., ‘dinosaur’ was tested amongst other animals), as were subordinate words (e.g., ‘Tyrannosaurus rex’ was tested in the context of other dinosaurs). The order of testing was randomised. See Appendix 1 for a summary of the words tested.

Procedure

Participants were presented with a series of 20 metaphorical statements constructed from the lists above, with supporting context. For example, 'This morning, Zork was playing outside in the garden. He picked up an apple and said, 'Look, let's play with this. The apple is a ball.' Half of the metaphors were presented at the basic level and the other half at the subordinate level, the selection of which was randomised for each participant. However, there were two instances where the same basic level vehicle occurred in two metaphors ('fish' and 'insect'), so there was a further control to ensure that participants were not presented with the same basic level vehicle twice. There were also 10 literally true and 10 literally false control statements. In total, there were 40 trials, the order of which was randomised for each participant. On each trial, the participant was required to rate the metaphor according to the 'teddy' system described above and was then asked to justify the response made. Both the response and the justification were dependent variables.

Prior to testing, the task was explained and comprehension of the scoring system was examined for one metaphorical statement, one literally true statement, and one literally false statement. Younger children were reminded of the scoring system throughout the experiment.

Results

Judgements

The data were analysed using a two-way mixed-design ANOVA, with Statement Type (*Basic, Subordinate, Literally True, Literally False*) as the within-subjects factor and Age Group (*4–5, 7–8, 9–10*) as the between-subjects factor. Descriptive statistics of judgement scores for each statement type are given in Table 4.

TABLE 4
Mean number of correct responses by age group and statement type (maximum = 5).

	Statement type							
	Basic		*Subordinate*		*Literally True*		*Literally False*	
Age group	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
4–5 years	2.63	0.79	2.03	0.75	4.73	0.31	1.43	0.42
7–8 years	2.48	0.73	2.69	0.79	4.99	0.31	1.26	0.36
9–10 years	2.57	0.73	3.09	0.52	4.93	0.16	1.13	0.24

There was a reliable main effect of Statement Type, $F(2.229,\ 156.054,$ Greenhouse-Geisser$) = 803.468$, $p < .001$, $\eta_p^2 = .920$, but not of Age Group, $F(2, 70) = 2.297$, $p = .108$, $\eta_p^2 = .062$. The interaction of Statement Type and Age Group was significant, $F(4.459,\ 156.054,$ Greenhouse-Geisser$) = 12.498$, $p < .001$, $\eta_p^2 = .263$, indicating that the pattern of responding was different for each age group (see Figure 4). Focusing on the experimental hypothesis, a second two-way ANOVA was conducted, with only the metaphorical statements considered: Statement Type (*Basic, Subordinate*) was the within-subjects factor and Age Group (*4–5, 7–8, 9–10*) was the between-subjects factor. The main effect of Statement Type was not reliable, $F(1, 70) = 0.369$, $p = .545$, $\eta_p^2 = .005$, but there was a significant main effect of Age Group, $F(2, 70) = 3.845$, $p < .05$, $\eta_p^2 = .099$, which was qualified by a reliable interaction of Age Group and Statement Type, $F(2, 70) = 25.571$, $p < .001$, $\eta_p^2 = .422$. Post-hoc analysis of simple effects showed that 4–5-year olds demonstrated a clear preference for basic metaphors over subordinate ones, $F(1, 27) = 18.730$, $p < .001$, $\eta_p^2 = .410$, but this preference was reversed for the 9–10-year olds, $F(1, 24) = 35.579$, $p < .001$, $\eta_p^2 = .597$; the 7–8-year-olds showed no reliable difference in their selection of each metaphor type, $F(1, 19) = 3.374$, $p = .082$, $\eta_p^2 = .151$. In addition, there was no reliable effect of Age Group with analysis restricted to basic metaphors, $F(2, 70) = 2.52$, $p = .778$, $\eta_p^2 = .007$, but a significant one when restricted to subordinate metaphors, $F(2, 70) = 16.248$, $p < .001$, $\eta_p^2 = .317$. Tukey HSD showed that 4–5-year-olds

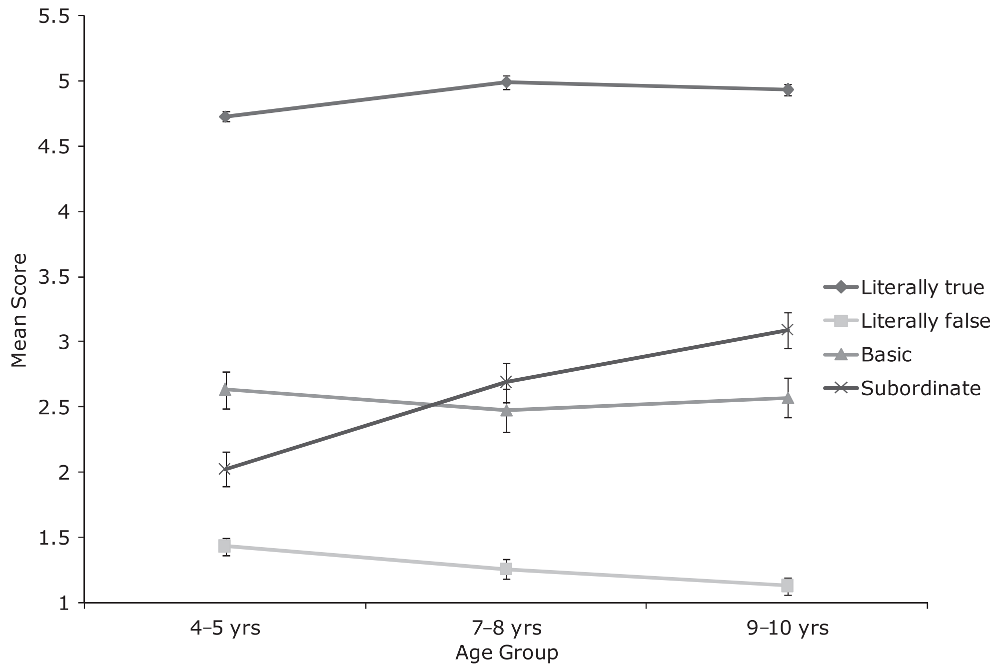

Figure 4. Mean judgement scores by age group and statement type. Error bars depict standard errors of the means.

had reliably lower preference scores for subordinate metaphors than both 7–8-year-olds, $p < .001$, and 9–10-year-olds, $p < .001$, but there was no significant difference between 7–8-year-olds and 9–10-year-olds on this measure, $p = .134$.

Justifications

Classification criteria and examples of justifications are given in Table 5. Interrater reliability was 0.80, based on 20% of responses. There were two coders, blind both to the age group of each participant and to experimental condition. The justification data were analysed with a three-way mixed-design ANOVA, with Statement Type (*Basic, Subordinate*) and Justification Type (*Basic, Subordinate, Literal, Semantic*) as the within-subjects factors and Age Group (*4–5, 7–8, 9–10*) as the between-subjects factor. Descriptive statistics of justifications are given in Table 6. There was no significant main effect of Statement Type, $F(1, 70) = 0.474$, $p = .493$, $\eta_p^2 = .007$. However, there were reliable main effects of Justification Type, F(1.852, 129.632, Greenhouse-Geisser) = 44.324, $p < .001$, $\eta_p^2 = .338$, and of Age Group, $F(2, 70) = 3.376$, $p < .05$, $\eta_p^2 = .088$. All two-way interactions were reliable ($p < .05$) and were qualified by the three-way interaction of Statement Type, Justification Type, and Age Group, F(4.174, 146.084, Greenhouse-Geisser) = 9.935, $p < .001$, $\eta_p^2 = .221$ (see Figure 5).

Focusing on the experimental hypothesis, there was a reliable effect of Age Group for *subordinate metaphor/subordinate justification*, $F(2, 72) = 31.326$, $p < .001$, $\eta_p^2 = .472$. Tukey HSD showed that 4–5-year-olds made significantly

TABLE 5
Classification criteria and example responses for justifications.

Classification criteria	*Definition of criteria*	*Example responses*
Literal	Metaphor taken at face value and interpreted as anomalous	'The apple isn't a ball/tennis ball.' 'It breaks, it's got pips.'
Semantic	Metaphor interpreted according to its contextual meaning, as if from a story	'Because he was outside.' 'He wanted to play.' 'He was naughty.'
Basic level attributes	Metaphor interpreted using attributes consistent with basic level objects	'It's like a ball, round and red.' 'You can pretend it's a ball because it's a circle.'
Subordinate level attributes	Metaphor interpreted using attributes consistent with subordinate level objects	'You can play with an apple, it's like a green tennis ball.' 'Apples are round – could play catch with them.'

TABLE 6
Mean frequency of justification types (see Table 5), by age group (maximum = 10).

	Justification type									
	Basic metaphor, Basic justification		*Subordinate metaphor, Subordinate justification*		*Basic metaphor, Subordinate justification*		*Subordinate metaphor, Basic justification*		*Non-metaphorical justifications (literal and semantic)*	
Age-group	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
4–5 years	3.75	3.17	2.89	2.71	0.11	0.31	0.14	0.36	7.57	4.61
7–8 years	5.35	2.21	6.75	1.89	0.55	0.60	0.05	0.22	3.85	3.31
9–10 years	5.44	2.40	7.60	2.08	0.52	0.77	0.00	0.00	4.04	2.81

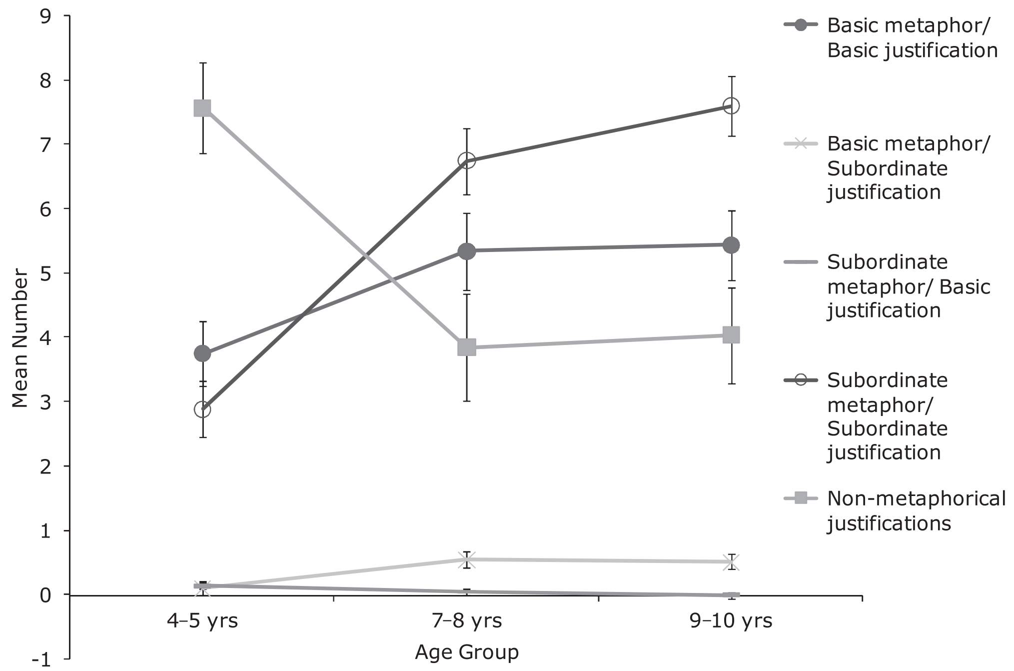

Figure 5. Mean number of justifications by age group and metaphor type. Error bars depict standard errors of the means.

fewer subordinate justifications for subordinate metaphors than both 7–8-year-olds, $p < .001$, and 9–10-year-olds, $p < .001$, but there was no significant difference between 7–8-year-olds and 9–10-year-olds, $p = .439$. There was no significant effect of Age Group for *subordinate metaphor/basic justification*, $F(2, 72) = 2.230$, $p = .115$, $\eta_p^2 = .060$.

There was a significant effect of Age Group for *basic metaphor/subordinate justification*, $F(2, 72) = 4.612$, $p < .05$, $\eta_p^2 = .116$. Tukey HSD showed that 4–5-year-olds made reliably fewer subordinate justifications for basic metaphors than both 7–8-year-olds, $p < .05$, and 9–10-year-olds, $p < .05$, but there was no significant difference between 7–8-year-olds and 9–10-year-olds, $p = .984$. There was also a significant effect of Age Group for *basic metaphor/basic justification*, $F(2, 72) = 3.283$, $p < .05$, $\eta_p^2 = .086$, but Tukey HSD revealed no reliable differences between the age groups, all $p > .05$. For brevity, the non-metaphorical justifications are considered together: there was a reliable effect of Age Group for such justifications, $F(2, 72) = 8.154$, $p < .001$, $\eta_p^2 = .189$, which Tukey HSD showed to be manifested in reliably more non-metaphorical justifications made by both 7–8-year-olds, $p < .01$, and 9–10-year-olds, $p < .01$, then by 4–5-year-olds, but no significant difference between 7–8-year-olds and 9–10-year-olds, $p = .984$. To conclude, Figure 4 shows a developmental rise in subordinate-level justifications for subordinate-level metaphors, with the significant effect arising between the 4–5-year-olds and the two older groups, accompanied by a corresponding fall in non-metaphorical justifications. The apparent rise in basic-level justifications for basic-level metaphors was not shown to be reliable by post-hoc tests.

GENERAL DISCUSSION

The main aim of the experiment was to test a novel prediction stemming from the MPC model of metaphor comprehension, namely that there is a developmental progression in the way that children process metaphors, from a preference for basic-level metaphors to a preference for subordinate-level metaphors. A judgement task was employed, in which children's preference for different kinds of similarity statements was assessed. In addition, participants were required to justify their preference judgements.

In line with the prediction from MPC, it was found that 4–5-year-old children preferred basic-level metaphors to subordinate-level ones, while 9–10-year-olds, conversely, preferred subordinate-level metaphors to basic-level ones. This change in preference was caused by a developmental increase in preference for subordinate level metaphors, since preference for basic metaphors remains relatively constant with increasing age. The results of the current study are therefore consistent with evidence concerning the development of categorisation (Mandler, 2007), suggesting that younger children organise conceptual information into basic-level categories based on large differences between concepts, with a shift towards subordinate-level categories with finer, more subtle distinctions, as children get older. The current results are also in line with previous suggestions that there is a relationship between metaphor development and conceptual categorisation abilities.

It should be pointed out that there was no significant decrease in preference for basic-level metaphors that corresponded to the increase in preference for subordinate-level ones. This may be because 9–10-year-old children are better generally at interpreting metaphors than the two younger age groups. Alternatively, it could be because, although the older children's conceptual knowledge would have become more refined and detailed than that of the younger children, the basic level conceptual categories are not, in fact, incorrect; each does not invalidate the other. One interpretation of a decrease in preference for basic-level metaphors would be that basic- and subordinate-level processes are dissociated from each other – the one replacing the other. The results of the current study, however, suggest that subordinate level cognition may instead be a development of basic level processes. One possible line of future research would be to test adult participants with the same paradigm. It is conceivable that adults would show a marked decrease in preference for basic level metaphors, compared to children, with the change occurring over a longer time scale than explored in this experiment. It is worth emphasising this point, since some of the basic-level metaphors used in the current study may not appear very apt to some readers. In any case, what is important for testing the prediction of the model is not the absolute rating of the metaphor types, but the difference between the two metaphor types across age ranges.

Beginning with the judgement scores, these show that, consistent with both developmental research (Vosniadou & Ortony, 1983) and the MPC model, 4–5-year-old children are able to distinguish metaphors from literally similar and anomalous statements. It is striking that, despite a small increase in the ratings given for the metaphors, they still remain relatively low. This may reflect the fact that a lack of wider meaningful context can make comparison statements relatively hard to interpret. In the current experiment, context was given by no more than a couple of introductory sentences. There is a difficult balance to be maintained, however, as the more context that is provided, the less metaphorical and the more literally similar the comparisons become. For example, if in the metaphor 'the dog is a Tyrannosaurus rex', 'dog' were changed to 'the dog that is big and scary and has a loud, frightening, roaring bark', comparison with a Tyrannosaurus rex might seem more literally than metaphorically similar. Conversely, too little context will make similarity statements increasingly anomalous. Thus, with no context at all, a dog is clearly *not* a Tyrannosaurus rex. As highlighted in the MPC account, discourse context is required to determine the status of a comparison that is ambiguous in isolation. A further point to note is that, while conclusions may be drawn about the different patterns of responding by the three age-groups, the fact that the contextual information provided by the experimenter was somewhat artificial prevents conclusions from being drawn about absolute ability levels at specific ages.

Children's justifications were analysed to investigate whether the attributes transferred from topic to vehicle shifted from basic- to subordinate-level across development. This line of evidence, too, indicated that there was a marked increase in the interpretation of subordinate level metaphors using subordinate level attributes. This increase was evident between the 4–5-year-olds and 7–8-year-olds, but there was no evidence that this increase continued with the 9–10-year-old children. In addition, there was a small but reliable increase in subordinate level justifications of *basic* level metaphors. This resulted from older children rejecting the basic level attributes supplied by the basic level metaphors, producing or even demanding more detailed information instead. These shifts further support the hypothesis that, as children get older, their semantic information becomes organised into subordinate-level categories containing increasingly detailed and subtle distinctions.

As with the judgement scores, there was also a reliable increase in the transfer of basic level attributes for basic level metaphor with increasing age, although no reliable differences were found between age groups with the particular post-hoc test used. It would likely represent a general improvement in the ability to comprehend metaphors as children get older, suggested by the fact that there is a marked decrease in non-metaphorical justifications given by the two older age groups. Improvements would be expected in, for example, children's ability to retain the information that is in the metaphors themselves. Since the justifications are measured in terms of a paraphrase or explanation, this requires both the child's comprehension, as well as their ability to recall, process and articulate what they have understood from the metaphor. Providing an interpretation of the metaphors therefore demands various cognitive skills that may be on different developmental trajectories. Any increase in basic level justifications of basic level metaphors would, therefore, partly reflect the child's developing abilities to understand, analyse and recall the information pertinent to each metaphor – something at which the 4–5-year-old children are likely to be less skilled. Clearly, this could also explain the increase in subordinate-level attributes transferred with age. However, the fact that there is a relatively larger increase in subordinate-level justifications suggests that, although children are generally better able at interpreting metaphors as they get older, this ability is more evident with subordinate level information. It seems reasonable to conclude, then, that the attributes transferred from topic to vehicle move from basic to subordinate level during development, consistent with the prediction of the MPC model.

The current study relates metaphorical thinking to the way in which conceptual information is organised. In line with the notion that concepts cluster together representations into coherent categories on the basis of similarity, by applying existing concepts to new experiences, one can infer meaning and learn from relatively unfamiliar situations. Since metaphors can also involve the transfer of information from a familiar category to a less familiar category, it has been argued that understanding metaphor makes use of mechanisms that are crucial for the acquisition of new information (Rumelhart & Norman, 1981). This notion has several implications for developmental theories of metaphor. Developmental research suggests that children relate new knowledge to existing concepts using analogical processes from a very early age (see Mandler, 1983). Metaphorical thinking in both adults and children may reflect this underlying shift of conceptual information from one domain to another, thus enabling children to apply existing knowledge to newly experienced phenomena. Metaphors therefore provide one of the fundamental ways of learning about the world. In light of this role, the shift in attributes transferred from the basic level to the subordinate level is unsurprising and may simply reflect the fact that, as children develop, more meaning is associated at the subordinate level and therefore more information is available to be linked to the topic of a metaphor.

The results of the current study, then, support the view that metaphor development is not a specific linguistic skill acquired at some point in development, but is related to developing conceptual processes. Moreover, in line with Mandler (2007), the results suggest that young children's conceptual knowledge is organised into categories or clusters and that these become increasingly delineated with advancing age. Furthermore, the results conform to the view that metaphorical thinking reflects the underlying transfer of information from one concept to another, a process that plays an important role in the child's acquisition of new knowledge.

Most importantly, perhaps, the current study illustrates how computational modelling may advance understanding of psychology (see Mareschal & Thomas, 2007). First, modelling advances theory development by demanding that researchers provide a formal specification of a given theory. Second, models can generate specific, novel, empirically testable predictions. In the case of MPC, such a prediction was tested and found to support the model. This success thereby provides evidence for the attendant assumptions of the model (Thomas & Mareschal, 1997, 2001), such as that the comprehension of novel metaphors is not achieved by a domain-specific mechanism but by the strategic use of semantic memory systems involved in categorisation, enabled by creative use of language. The empirical confirmation is also consistent with the assumption that similarity is a transformational process, although it may be possible to derive similar predictions from models with different assumptions, such as feature-based models. This is an issue that could be addressed in future work. Another possible area for future investigation would be to see whether MPC could be adapted to include relational similarity. Although it is unclear whether such an extension would be possible, Leech et al. (2008) have demonstrated how relations could be encoded for a connectionist network. Although clearly very simple, MPC has successfully captured various phenomena here; it may be extendable to more.

Manuscript received September 2008
Revised manuscript received January 2009
First published online month/year

REFERENCES

Billow, R. M. (1981). Observing spontaneous metaphor in children. *Journal of Experimental Child Psychology*, *31*, 430–445.

Black, M. (1979). More about metaphor. In A. Ortony (Ed.), *Metaphor and thought*. Cambridge, UK: Cambridge University Press.

Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*, 193–216.

Chandler, S. R. (1991). Metaphor comprehension: A connectionist approach to implications for the mental lexicon. *Metaphor and Symbolic Activity*, *6*, 227–258.

Chukovsky, K. (1968). *From two to five*. Translated and edited by Miriam Morton. Berkeley, CA: University of California Press.

Gardner, H., Winner, E., Bechhofer, R., & Wolf, D. (1978). The development of figurative language. In K. Nelson (Ed.), *Children's language: Volume I*. New York: Gardner Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge, UK: Cambridge University Press.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, *97*, 3–18.

Goswami, U. (1996). Analogical reasoning and cognitive development. In H. Reese (Ed.), *Advances in child development and behaviour*. New York: Academic Press.

Gottfried, G. M. (1997). Using metaphors as modifiers: Children's production of metaphoric compounds. *Journal of Child Language*, *24*, 567–601.

Hahn, U., Chater, N., & Richardson, L. B. C. (2003). Similarity as transformation. *Cognition*, *87*, 1–32.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*, 295–355.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. New York: Harper & Row.

Keil, F. C. (1986). Conceptual domains and the acquisition of metaphor. *Cognitive Development*, *1*, 73–96.

Leech, R., Mareschal, D., & Cooper, R. (2008). Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences*, *31*, 357–378.

Mandler, J. M. (1983). Representation. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology*. New York: John Wiley.

Mandler, J. M. (2007). On the origins of the conceptual system. *American Psychologist*, *62*, 741–751.

Mareschal, D., & Thomas, M. S. C. (2007). Computational modeling in developmental psychology. *IEEE Transactions on Evolutionary Computation (Special Issue on Autonomous Mental Development)*, *11*, 137–150.

Marschark, M., & Nall, L. (1985). Metaphoric competence in cognitive and language development. In H. Reese (Ed.), *Advances in child development and behavior* (Vol. 19). New York: Academic Press.

McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In J. L. McClelland, D. E. Rumelhart (Eds.), & The PDP Research Group , *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models* (pp. 170–215). Cambridge, MA: MIT Press.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50A*, 528–559.

Morrison, C. M., & Ellis, A. W. (1995). Roles of frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 116–133.

Piaget, J. (1962). *Play, dreams and imitation in childhood*. New York: Norton.

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*, 451–469.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Rumelhart, D. E., & Norman, D. A. (1981). Analogical processes in learning. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Sun, R. A. (1995). A microfeature based approach towards metaphor interpretation. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 424–429). San Francisco, California: Morgan Kaufmann.
Thomas, M. S. C., & Mareschal, D. (1997). Connectionism and psychological notions of similarity. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 757–762). London: London Education Authority.
Thomas, M. S. C., & Mareschal, D. (2001). Metaphor as categorisation: A connectionist implementation. *Metaphor and Symbol*, *16*, 5–27.
Vosniadou, S. (1987). Contextual and linguistic factors in children's comprehension of non-literal language. *Metaphor and Symbolic Activity*, *2*, 1–11.
Vosniadou, S. (1989). Context and the development of metaphor comprehension. *Metaphor and Symbolic Activity*, *4*, 159–172.
Vosniadou, S., & Ortony, A. (1983). The emergence of the literal-metaphorical-anomalous distinction in young children. *Child Development*, *54*, 154–161.
Winner, E. (1979). New names for old things: The emergence of metaphoric language. *Journal of Child Language*, *6*, 469–491.

APPENDIX 1

Words used in vocabulary test:

Basic level

Insect, bug, dinosaur, broom, hair, jewels, instrument, bear, necklace, building, stick, coin, tyres, pasta, sponge, cloud, chimney, ape, rain, fruit.

Subordinate level

Dolphin, starfish, ants, stick insect, bee, tennis ball, Tyrannosaurus rex, eagle, fringe, diamond, flute, teddy bear, Smarties, racing car, tower, twigs, ten pence piece, trainers, spaghetti, raincloud, top hat, gorilla, raindrops, matchbox.