



## Children's humour development: A linguistic perspective

Giulia Baker<sup>1</sup>

*Freelance researcher & Teacher*

Michelle Aldridge<sup>2</sup>

*Cardiff University*

Received : 20.08.2022  
Accepted : 12.11.2022  
Published : 30.11.2022

### Abstract

By applying findings from an investigation into English-speaking children's ability to comprehend verbal riddles, this paper advances an established model of children's humour development by expanding stage 5 to include four sub-stages: (a) lexical and phonological ambiguities; (b) morphological ambiguities; (c) syntactic ambiguities and (d) idiomatic ambiguities. Sixty children equally divided from three British school Year Groups: Year 2 (aged 6-7), Year 4 (8-9) and Year 6 (10-11) participated. Their understanding of riddles was measured (a) receptively through a multiple-choice task in which they were required to identify an ambiguous punchline and (b) productively through a verbal explanation task in which they were required to explain their understanding of a riddle containing an ambiguous word/phrase. Responses were analyzed using a linguistically based classification system and explanations are offered as to why some ambiguity types are easier/harder for young children to comprehend.

**Keywords:** child humour, verbal riddles, ambiguities, children's cognitive and social development, literacy development

### 1. Introduction

Studies on children's humour development, especially within the field of linguistics, have been sparse in recent decades despite the body of evidence to suggest that humour is effective in developing creativity and divergent thinking (Ziv, 1976, 1983, 1988), developing higher order reading skills (Yuill, 1998; Zipke, 2007; 2009), raising reading levels (Yuill, 1998; Zipke, 2008) and developing confidence in social communication and interactions (Graham, Papa & Brooks 1992, Nezlek & Derks 2001). Humour is also known to be a motivational factor in literacy development, especially in the riddle form as it is perceived as fun, enjoyable and the joking format most favoured and recognised by young children (Wolfstein, 1954; Zipke, 2007, 2008).

The shortage of recent research on children's verbal humour development from within the field of linguistics is somewhat paradoxical given that verbal humour is intrinsically based within the language through which it is

<sup>1</sup> Giulia Baker is a researcher, teacher and former Literacy Coordinator who has taught across all School Year Groups in the UK, as well as at undergraduate level. Her interests lie in the ways different types of ambiguity-based humour can be used in the classroom to motivate pupils and facilitate learning.

<sup>2</sup> Michelle Aldridge is a Reader at Cardiff University, she lectures in psycholinguistics & forensic linguistics. Her research focusses on the communicative experiences of vulnerable people including children and adults with a communication disorder. She writes widely in journals such as cognitive linguistics, forensic linguistics, health and communication and language awareness. Corresponding author: [AldridgeM@cardiff.ac.uk](mailto:AldridgeM@cardiff.ac.uk)

communicated and arises from the manipulation of different linguistic structures. Here, we investigate the gap in both the literature and our understanding about the way in which different linguistic phenomena affect children's humour comprehension through a robust linguistically based classification system (Baker and Aldridge 2021) and, in so doing, the originality of this paper lies in advancing our theoretical understanding of children's humour development, together with progressing the final stage of a recognized model of humour development (McGee 1979/2002).

Given our linguistic focus, we acknowledge the work achieved through Raskin's (1985) Semantic Script Theory of Humour (SSTH), later developed with Attardo (1991) as the General Theory of Verbal Humour (GTVH) to facilitate categorization of all verbally expressed jokes. The jokes (riddles) we employ in this study are verbally expressed but are also specific in that they rely upon the manipulation of linguistic phenomena for their humour. This type of verbal joke must be 'understood' or made sense of in order that any humorous response be activated. We therefore adopt the Incongruity Resolution (IR) theory of humour (Suls 1972, 1983) as our framework because it is based upon understanding, 'the intellectual part of the humour reaction which is an indispensable basis for amusement' (Bariaud 1989, p. 20).

Within this framework, we adopt McGhee's humour model which, although formulated in 1979 and revised in 2002, still comprises the established framework for children's humour development. We focus upon McGhee's final stage which typically represents humour development around seven years of age as we are interested in capturing children's understanding of verbal incongruities, multiple meanings, and ambiguities necessary for processing verbal riddles (Bariaud, 1989; McGhee, 1977, 1979, 2002; Shultz, 1974; Shultz & Horibe, 1974). To contextualize this stage, we outline, in Table 1, McGhee's model of humour development which comprises all five stages through which children sequentially progress.

Table 1

*Summary of McGhee's Revised Stages of Humour Development (2002)*

Stage 0: Laughter without humour (0 to 6 months)
Stage 1: Laughter at the attachment figure (6 to 12-15 months)
Stage 2: Incongruous actions towards objects (12-15 months to 3-5 years)
Stage 3: Incongruous labelling of objects or actions (2 to 3-4/5 years)
Stage 4: Conceptual incongruity (3 to 5 years)
Pre riddle stage: transition period (5 to 6 -7 years)
Stage 5: Riddles and jokes (double meanings): 6-7 to 10-11 years
<ul style="list-style-type: none"> <li>• Multiple meanings: ambiguity (in puns); first signs of logic (in riddles)</li> </ul>



“Understanding” humour during the final stage (Five) of this model requires complex processing as the listener must think about the linguistic components of the riddle and be aware of their impact on the listener. The development of metalinguistic skills (cf. Tunmer, Pratt & Herriman 1984) indicates that the child is not only using language to communicate but must also know what they are using and to what end (e.g., Hoppe & Kess 1982). More specifically then, in these riddles participants must identify that some linguistic unit has two (or more) meanings; that is, they must locate an incongruity (i.e., an ambiguous word or phrase) and be able to go back and forth between these meanings to determine how they both make sense within a single context (i.e., they ‘resolve’ - or make sense of - the incongruity). Dissecting a humorous text in this way, requires the listener to work harder in terms of cognitive and linguistic processing than does a non-ambiguous text and it is argued (e.g. Bariaud, 1989; Fowles & Glanz, 1977; McGhee, 1971a, 1972, 1977, 1979, 2002) that children are typically seven years of age before they start to develop the requisite cognitive skills to process such ambiguities and thus transition from Stage Four to Stage Five (McGhee, 1979, 2002). Associated with this ability is, of course, the development of pragmatics and social cognition (Hyter, 2017), that is, at very least, the listener must have a theory of mind to appreciate that the speaker is engaging in humour, deviating from the fundamental principle that speakers are cooperative in communication (Grice, 1989) and able to engage in fantasy assimilation (McGhee, 1979).

In addition to evolving cognitive, social cognition and metalinguistic skills, children’s general language development also influences humour advancement. Specifically, in order to progress to the final humour stage, children must understand the ways in which sounds, words, morphology and syntax operate together to communicate meaning, knowledge which is acquired during the pre-school years (Berger, 2011; Hoff, 2015; Kaplan, 1998) and continues throughout the primary years as they acquire later aspects of linguistic proficiency, such as idiomatic language, (Ackerman, 1982; Cain, Towse & Knight, 2009; Gibbs, 1987; Levorato & Cacciari, 1995; Prinz, 1983).

Although acquisition of the above skills is rapid during pre and early school years, verbal humour is not fully understood during this period (McGhee 1979, 2002; Bariaud 1989). Whilst children may be able to understand jokes expressed through the medium of language (‘verbally expressed’ humour), they are not yet able to understand jokes reliant upon the manipulation of linguistic properties (‘verbal humour’) which rely upon punning, double entendres and ambiguities (e.g., ‘Why can’t a car play football?’ ‘Because it’s only got one boot.’). This type of understanding only starts to develop as they mature cognitively and linguistically and enter the final humour stage (McGhee 1979, 2002). Even then, not all ambiguity-based humour is understood with the same facility since ambiguity is manifested at varying levels of language and requires different types of linguistic proficiency to be understood. Some ambiguity types are reported to be easier to comprehend than others (Shultz & Pilon, 1973; Shultz, 1974; Shultz & Horibe, 1974; Hirsh-Pasek, Gleitman & Gleitman, 1978; Yuill, 1998) but earlier research findings on humour comprehension are inconclusive, resulting in little

consensus as to which types of ambiguities are most readily understood by young children.

Part of the prevailing lack of consensus may be because previous research has often been carried out in the field of psychology (as opposed to linguistics) where different types of language-specific ambiguities were included with emphasis centered upon developing a framework to account for children's humour development, rather than upon the properties of language being tested. Earlier studies frequently lack ambiguity definitions, which leads us to question which language phenomena they embody. Even when included, definitions, as described below, are often broad and generic and typically not interpreted or applied in the same way. This has resulted in 'the same linguistic phenomena being tested under different ambiguity classifications and different linguistic phenomena being tested under the same ambiguity classification' (Baker & Aldridge 2021, p.239).

Taking, for example, lexical ambiguity, this ambiguity has often been used in studies on children's humour/ambiguity comprehension but historically interpreted in different ways both within and across studies: Shultz and Pilon (1973, p.728) describe it as occurring 'when a given lexical item has more than one semantic interpretation', classifying 'bank/ bank' (river or financial) as lexical ambiguity but place 'pears/pairs' in the category of phonological ambiguity (despite their containing no sound modification when orally represented). Similarly, Shultz and Horibe (1974) detail phonological ambiguity as occurring 'when a given phonological sequence can be interpreted in more than one way' (Shultz & Horibe, 1974, p.14) when only one of their examples relies upon sound modification (line/lion), the other two relying instead upon homophony (pear/pair) and mis-parsing (eighty cups/eight teacups). Such inconsistencies have resulted in different linguistic phenomena being tested within the same category of ambiguity which means reported findings do not necessarily reflect outcomes for the linguistic phenomena they purport to test e.g., Shultz and Pilon (1973) report that children find phonological ambiguities easiest to detect even though some of their stimuli rely upon homonymy and the shifting of word boundaries rather than upon sound distortion and the manipulation of phonemes. This in turn casts doubt upon the validity of findings, especially as ambiguities manifested through manipulation of different linguistic phenomena require different processing skills and pose differing levels of challenge. To overcome these limitations our data are classified according to our linguistic classification system (Baker and Aldridge 2021).

Of particular note is the tight link between the lexicon and syntax and the categorization of riddles such as 'Why do leopards make rubbish thieves? Because they're always spotted.' Understanding this riddle requires identification of a lexical ambiguity located in the homonym 'spotted' but this is not enough per se. Although the two homophones have identical surface representations, spotted ('covered with spots') and spotted ('was seen'), they nonetheless have two different syntactical representations (an adjective/past tense verb form) which means this ambiguity requires grammatical processing at a deeper level. To make sense of the ambiguity the listener must identify not only the homophone 'spotted' but differences in word class too. Zipke (2007, p.382) acknowledges this difference by restricting lexical



ambiguity to cases where ‘a word has more than one meaning without a class violation,’ but others do not - Yuill (1998), for example, classifies the ‘spotted riddle’ as comprising lexical ambiguity – rather than syntactic – and does not accommodate the fact this ambiguity requires additional processing skills in order it be understood.

Other - higher order - processing skills, such as the decoding of figurative language, have also been overlooked in earlier studies, despite ambiguities based upon this linguistic phenomenon having been included. Researchers do not always accommodate the fact that idiomatic ambiguities require a specific type of processing skill - that of figurative language - and instead include stimuli based upon this ambiguity type in different types of categories altogether, Fowles and Glanz (1997, p.446) categorise, ‘Why didn’t the skeleton cross the road? It didn’t have the guts’ as lexical ambiguity and Shultz and Pilon (1973, p.30) treat ‘He stepped over the line/lion’ as being phonologically ambiguous. Neither address the fact that idiomatic ambiguity requires linguistic knowledge which cannot be gleaned from the sum of its individual lexical items. This is a matter which we accommodate within the present study.

Differences in the interpretation and application of discrete ambiguity types as outlined above, has meant that findings regarding children’s humour comprehension are often inconsistent (Shultz & Pilon, 1973; Shultz and Horibe, 1974; Hirsh-Pasek et al., 1978; Yuill 1998) and this has consequently contributed to stagnation in our theoretical understanding of the final stage of humour development. As stated, our study addresses this matter by using Baker and Aldridge’s (2021) robust linguistically based classification system (see appendix A), to develop both the humour model and our understanding of the way in which children’s humour develops. We use these definitions because they were specifically established to improve classification of ambiguity types for practical investigation. They accommodate the fact that verbal humour is intrinsically embedded within the form in which it is delivered (i.e., the language through which it is communicated) and identify the specific linguistic phenomena through which discrete ambiguity types are manifested. Using these definitions allows us to be sure of the specific linguistic phenomena being tested at any given time – and of the processing skills required in order that any ambiguity be understood.

The three aims of this investigation were: (a) to establish how children’s comprehension of orally delivered verbal riddles differs across the ages of six to eleven years; (b) to determine whether ambiguity type affects the facility with which verbal riddles are understood by these children and (c) to establish whether our data can be explained within McGhee’s model of humour development. To this end we use verbal riddles, each reliant upon one of five ambiguity types (lexical, phonological, morphological, syntactic, and idiomatic) to test children’s comprehension of humour.

We hypothesis: (1) comprehension of ambiguity types will improve across ascending Year Groups, (2) the different ways in which linguistic properties are manipulated to create humour will affect participants’ understanding, and (3) ambiguities requiring more complex language processing skills will be less readily understood.

## **2. Methodology**

### *2.1. Participants*

Participants comprised sixty children from a mainstream English medium primary school with twenty children taking part from school Year 2 (aged six to seven years), Year 4 (aged eight to nine years) and Year 6 (aged ten to eleven years). These participants were selected because they spanned the transition period between Stage Four and Five of McGhee's (2002) humour model when verbal ambiguities are said to be first comprehended. Thirty-five of the participants were male and twenty-five were female. Gender distribution was evenly spread across Year Groups. Participants were all L1 English monolingual speakers and had no known sensory, learning, or behavioural disability.

### *2.2. Ethics*

Consent was granted by the Headteacher of the participating school and by parents/guardians of every participant (each of whom was provided with a letter detailing the nature and purpose of study and how data were to be collected, stored, and used). All data were anonymized. The University's ethics committee also approved the research.

Throughout the study the utmost care was taken to ensure that the participants did not experience discomfort or find the activities stressful either physically or psychologically. Irrespective of consent, if the researcher felt that a child was feeling upset or uncomfortable at any stage, the study was stopped immediately, and the child's participation terminated in a sensitive manner (this happened on a single occasion during the pilot study). Most participants appeared to find the activity enjoyable and they frequently expressed regret about it ending.

### *2.3. Data collection and processing*

A single researcher collected the data. This allowed for consistency in the oral delivery of the stimuli and numbers and types of prompts. The researcher was not known to the participants. Although a class teacher might have provided a more familiar face for administration of the study, this was not feasible since the process was considerably time-consuming and the teacher was occupied within the classroom. The researcher was an experienced practicing primary teacher, used to teaching and engaging with children across all participating Year Groups. Two additional coders were employed to rate the participants' responses (explanation scores) independently after the data had been collected.

The study took place during the school day and within the school building, a familiar environment for participants. Participants were withdrawn from their classrooms and taken to an alternative room where they were able to provide individual responses without being influenced by their peers and could be recorded without interruption. It took an average of fifteen minutes per child to complete the activity.

Verbal riddles (short question-answer jokes) (see appendix B) were used to test participants' ability to identify and explain different ambiguity types. Each riddle contained a single verbal ambiguity (a word or phrase that could



be interpreted in more than one way) in its punchline<sup>3</sup> aimed at eliciting a humorous response. Ambiguities were all either reliant upon words and meanings, being phonetically similar to some other string not present in the utterance or else relied upon two identically sounding utterances being lexically analyzable in two different ways. Pre-rating of ambiguities was piloted through classification by undergraduate Linguistics students and only riddles that met with full consensus were included. Each riddle contained more than a single word punchline answer, was of a similar length and was age appropriate (non-offensive and conceptually familiar to participants). Three riddles were used for each ambiguity type to reduce the possibility of participants selecting a correct punchline by chance. Processing demands and attentional abilities contributed to a limit of 15 riddles, three per ambiguity type for each participant. All riddles were trialled in a pilot study with participants of the same age to ensure that language/concepts were familiar to participants across Year Groups so that these factors would not impact upon processing skills.

#### 2.4. *Data analysis*

Comprehension was the focus of the investigation, and one of the challenging issues to address was how best to determine whether participants had understood a riddle based on ambiguous use of language. It was important that opportunities be maximised for participants to communicate their understanding to the researcher. Previous studies have addressed this issue through a variety of ways including using multiple choice exercises (Yalisove, 1978; McGhee & Panoutsopoulou, 1990; Yuill, 1998; Zipke, 2007), participants' recall (Yalisove, 1978; Fowles & Glanz, 1977; Yuill, 1998), participants' explanations (McGhee, 1971b; Shultz, 1974; Shultz & Horibe, 1974; Prentice & Fathom, 1975; Fowles & Glanz, 1977; Hirsh-Pasek et al, 1978) and graded comprehension scores (McGhee, 1971b; Prentice & Fathom, 1975; Fowles & Glanz, 1977; Hirsh-Pasek et al., 1978). Each of these methods was considered.

Recall was eliminated as a comprehension measure on the basis that many young children can recall and recite short texts verbatim (in this instance riddle punchlines consisting of three to nine words) in so called "parrot fashion", often without any real processing of the understanding of their underlying meanings (Allington & Strange, 1979). This left two other main options, multiple choice exercises and eliciting verbal explanations (which here incorporated graded comprehension scores). Both these methods were trialled in a pilot study to determine their efficacy as summarized in the next section.

#### 2.5. *Pilot study*

The multiple-choice task involved nine participants aged six to eleven years listening to a verbal riddle and then selecting one of three potential punchlines, only one of which was the correct (original) punchline. The verbal explanation task involved the same nine participants aged six to eleven years listening to a verbal riddle and then explaining how this

---

<sup>3</sup> Ambiguities in questions are reported as being more difficult to identify (Yalisove, 1978).

coupling made sense to them within a humorous context. These nine participants did not participate in the main study.

Analyses of results from the pilot study showed that comprehension of ambiguities was measured in both tasks but in distinct ways. The multiple-choice task tested comprehension in terms of being able to IDENTIFY an ambiguity while the verbal explanation task tested comprehension in terms of being able to EXPLAIN an ambiguity. Both required metalinguistic application but placed different demands on participants depending upon whether comprehension was being measured receptively (identification of an ambiguity) or productively (explanation of an ambiguity). The multiple-choice task alleviated language production demands relating to justification of punchline selections but did not give any insight into reasons behind punchline selections. The verbal explanation task, on the other hand, afforded an insight into reasons behind punchline selections, but was potentially more difficult for participants. It was therefore decided to combine the two tasks in the main study. Each provided a counterbalance to potential weakness in the other and by combining the two tasks, a richer set of data could be obtained than had only one of the tasks been carried out. Both ambiguity identification and explanation scores were treated as being indicative of comprehension of ambiguities in punchlines, but differences between the two tasks, the cognitive demands they placed upon participants, and their respective strengths and limitations were acknowledged.

### *2.6. Main study procedure*

Participants were tested individually by the researcher. Those who had a joke to share with the researcher (as suggested in the initial letter requesting consent for participation) related their joke as an initial icebreaker. The researcher then discussed what constitutes a riddle and explained that she was interested in finding out which types of riddles were understood by children in different school Year Groups. It was emphasized that participants were not taking part to give a "right" or "wrong" response but rather were sharing their ideas and understanding about individual riddles as a representative of their particular Year Group. A warm-up activity was carried out during which the researcher played the role of participant and chose a punchline from a selection of three and explained how the punchline made sense to her by identifying an ambiguous word and then explaining the two possible interpretations. Each participant was then given a practice try themselves before the task began.

The researcher discussed the fact that she might well use the phrase "And anything else?" as a prompt in the context of reminding participants that there might be an additional meaning to be explained but stressed that participants were not to feel pressurised if they had nothing further to add. She reassured the children that lots of people had nothing to say when given this prompt and that was acceptable. Participants were able to ask questions before the study began and throughout its duration. Whilst a few participants asked for clarification regarding procedures, most questions asked related to the technology used to record participants' explanations.

When the researcher was sure that each participant understood the activity through appropriately responding to the practice riddles, she read out fifteen





riddles, one by one, each with three potential punchlines (see Appendix B). Of the three potential punchlines, one was the riddle's correct (original) punchline containing an ambiguity upon which the riddle depended, one was a "plausible" punchline, and one was an "irrelevant" punchline. The "plausible" punchline treated the riddle question as if it were a bona fide request for information rather than one intended to elicit a humorous response and contained a logical answer to the riddle question (e.g., 'How did the banana know he was ill' PLAUSIBLE PUNCHLINE = 'He had a high temperature'). The "irrelevant" punchline was one which neither treated the question as a bona fide request for information, nor as a rejoinder intended to elicit humour (e.g., 'How did the banana know he was ill' IRRELEVANT PUNCHLINE = 'He looked out of the window'.) Both original punchline positions and ambiguity types were presented in random order although the order was identical for each participant within each riddle multiple choice selection. Participants were able to hear the riddle question and any of the punchlines as many times as desired. Participants then chose the punchlines which, when coupled with the riddles' interrogatives, they felt completed the riddle. Distribution of ambiguity types was spread across the activity to address any potential fatigue effect for riddles presented later (although subsequent analysis indicated no significant fatigue effect). Participants were asked to indicate if they had heard any of the riddles before. In such instances, an alternative riddle was supplied (again with a choice of three punchlines) based on the same ambiguity type and which met all the other criteria for inclusion<sup>4</sup>.

After choosing a punchline, participants were recorded explaining the reason for each punchline selection. When applicable, non-leading prompts were used e.g. "And anything else?", "Can you explain what you mean by . . .?" to obtain as full a response as possible. Upon completion of the two tasks participants were thanked for their help and participation.

### 2.7. Scoring

Identification scores were totalled for each type of punchline selection and for each ambiguity type correctly identified. Chi square tests and analyses of variance were run to determine whether differences in punchline selections by Year Group and/or ambiguity type were statistically significant.

Only explanations relating to correct punchline selections were scored since these directly related to the three main research aims although, explanations for non-target (i.e., plausible, and irrelevant) selections were also analyzed qualitatively to determine reasons why participants chose answers that were non-target non-original punchlines.

Three independent raters used comprehension criteria (cf. Comprehension Criteria) to score transcribed explanations. Homophones were transcribed phonetically so as not to influence raters. Morphologically ambiguous words

---

<sup>4</sup> The highest number of substitute riddles were for lexical ambiguity (22). This did not negatively impact upon comprehension since lexical ambiguities were identified most frequently overall in the multiple-choice and scored second highest in the verbal explanation test (cf. results) – which corresponds with findings from studies in which this type of ambiguity has been interpreted in a similar way.

were transcribed as either one or two words according to the juncture used by participants. Words that were notably stressed were underscored. Gestures that were made in support of, or instead of, participants' explanations were annotated in transcriptions.

Scores ranged from 0-2 depending on whether 1, 2 or 0 interpretations of the linguistic word/string were judged to have been communicated. When raters could not reach a consensus, scores were awarded according to the most frequently provided score (on no occasion did raters award three different scores). Scores were totalled for Year Groups and for ambiguity types. Analyses of variance were run to determine whether differences between Year Groups and/or ambiguity types were statistically significant.

Comprehension criteria accommodated all strategies employed by participants to communicate understanding. It did not matter which strategy participants used as long as it showed understanding. Meanings of words and phrases were judged to have been understood and communicated if a participant:

- 1 provided a definition/explanation of a meaning e.g. "Because um when they say seven /eit/ nine its um seven starts eating other numbers but /eit/ is also the number after seven, so they said seven /eit/ nine."
- 2 used context to illustrate understanding e.g. "Because /ʃi:təz/ cheat . . . because it means if I had snakes and ladders, I already had a go I would roll it. I would probably just say 'Can I have the dice for a sec?' turn around, get it to number six and go 'Yay, I got a six.'"
- 3 manipulated juncture to highlight differences in sound(s) e.g. "Because it's a milkshake you shake it and then the um oh watch- um the blender and it makes like um the milk the milk shake and the jelly usually wobbles and he saw the milk shake so he wobbled."
- 4 made contrastive reference to substituted phonemes e.g., "it's 'ch' instead of 'sh'"
- 5 manipulated a root word, either inflectionally or derivationally, to exemplify meaning e.g., used "cheat", "cheats", "cheating", "cheated" to illustrate meaning of the word "cheater": "In card games you have /ʃi:təz/ cheating. Somebody cheats."
- 6 deliberately articulated individual words so that they varied notably from conventional stress patterns i.e., used word stress to draw attention to meanings: "Because it was meant to be feeling well it's a banana peeling."
- 7 identified word clusters as containing meanings not readily analysable from the sum of individual parts e.g., referred to idioms as 'sayings', 'terms' and 'phrases': "Because flying off the handle is a phrase and the witch flies on a broom and she flies off the handle."
- 8 used a deictic or iconic gesture e.g., either pointed to something or made a movement typically associated with a word's meaning.
- 9 gave an answer that indirectly showed comprehension of one or more meanings of an ambiguous word/phrase e.g. "I chose that one because chickens can go like /ʃi:p ʃi:p/ and a /ʃi:p/ chicken would be quite good for Christmas dinner . . . because you don't, cos then you have more money to spend on presents."



Once these criteria had been applied to score explanations, results were analysed as summarised below.

### 3. Findings

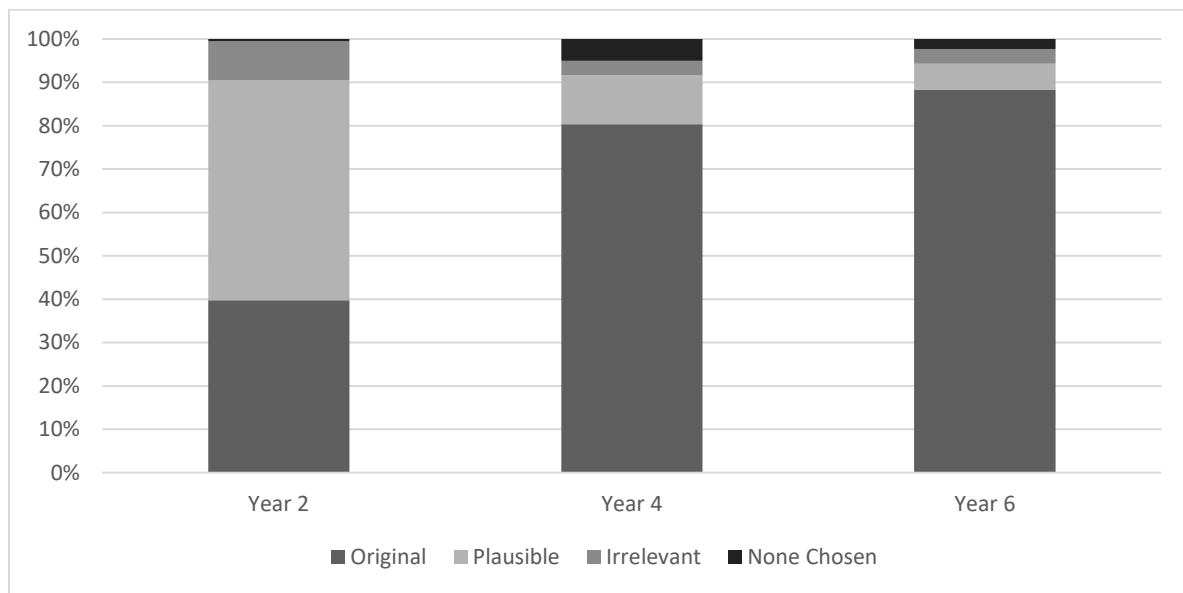
#### 3.1. Findings according to year group

As shown in Table 2, none of the Year Groups correctly selected the target (i.e., original) punchline for all fifteen riddles. More specifically, no participant in Year 2 selected all fifteen original punchlines correctly and only two participants in Year 4 (10%) and four participants in Year 6 (20%) chose all fifteen correct (original) punchlines. Using the binomial distribution ( $p=1/3$ ,  $n=300$ ,  $x=147$ ) with Original punchlines as markers of success, the probability of Year 2 choosing 147 Original punchlines by chance is  $<0.0001\%$ . For Years 4 and 6, whose Original punchline choices were greater, the probability is even less. We can therefore be confident that original punchline selections were not attributable to chance.

Table 2  
*Punchline Selections Made by Year Groups in the Multiple-Choice Task*

	Year 2	Year 4	Year 6	Total number of punchlines chosen per category across all Year Groups
<b>Original Punchline Selected</b>	147 (49%)	241 (80%)	265 (88%)	653
<b>Plausible Punchline Selected</b>	118 (39%)	34 (11%)	18 (6%)	170
<b>Irrelevant Punchline Selected</b>	33 (11%)	10 (3%)	10 (3%)	53
<b>No Punchline Selected</b>	2 (1%)	15 (6%)	7 (3%)	24

A chi square test run was run on all multiple-choice selections to determine whether there was a statistically significant difference in selection choices made by individual Year Groups. The chi square test showed there to be a statistically significant (increased) difference as Year Groups ascended: ( $\chi^2(4) = 155.4$ ,  $p < 0.01$ ). This increase was expected given participants' general cognitive stages of development. There was, however, a marked difference in the way in which this increase evolved between Years 2 (aged six to seven) and 4 (aged eight to nine) and Years 4 (aged eight to nine) and 6 (aged ten to eleven) as illustrated in Figure 1.



*Figure 1.* Punchline selections made by individual year groups in the multiple-choice task

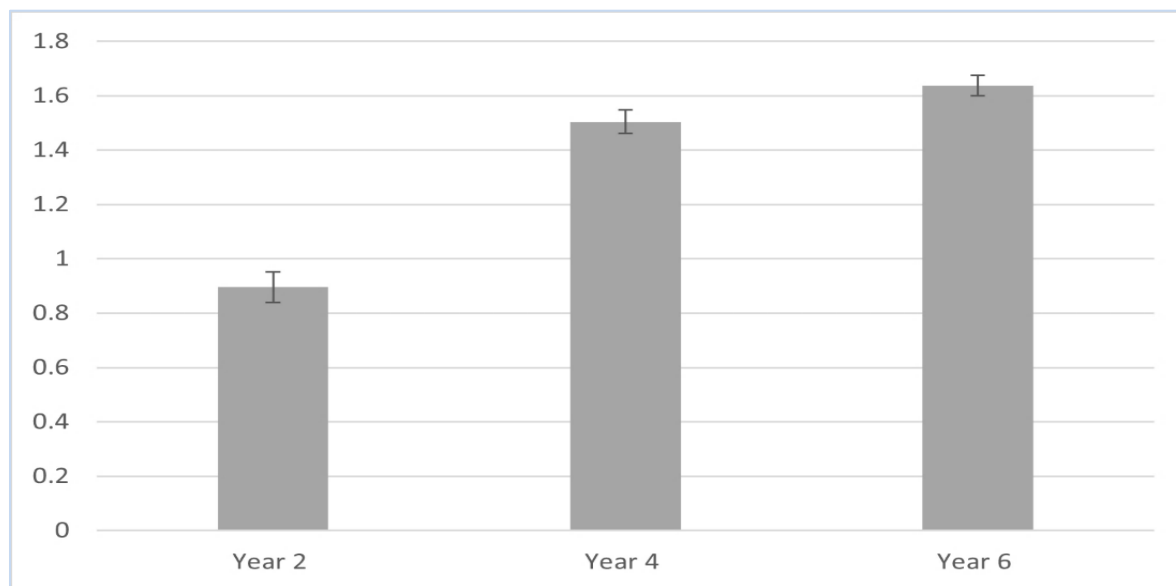
Although a significant difference in the ability to identify correct (original) punchlines existed across the three Year Groups, this difference was not equally spread. There was a greater increase in correct punchline choices between Years 2 and 4 than between Years 4 and 6 (the differences were 94 and 24 respectively). Two separate post-hoc chi square tests were performed, one on the difference between multiple choice selections for Years 2 and 4 ( $\chi^2(2) = 81.2, p < 0.01$ ) and one on the difference between Years 4 and 6 ( $\chi^2(2) = 6.0, p > 0.05$ ). The difference between correct punchline selections was statistically significant for Years 2 (aged six to seven) and 4 (aged eight to nine) (147 correct choices as opposed to 241) but not for Years 4 (8-9) and 6 (10-11) (241 correct choices as opposed to 265). The difference in correct (original) punchline selections was statistically significant only between the two youngest Year Groups. This would therefore suggest a period of accelerated development in the ability to identify ambiguity-based punchlines to occur somewhere between the ages of six to nine years and we will return to this point in the discussion.

Only explanations relating to correct punchline selections were scored, although explanations for non-target (i.e., plausible, and irrelevant) explanations were also analysed qualitatively to determine reasons why participants had chosen non-target punchlines. Using the comprehension criteria as detailed above, but scoring individually, three independent raters agreed with 90.6% of the explanations scored. Following discussion, they agreed on 98.6% of scores. For the nine explanations upon which raters did not all agree, there was always consensus between two of the three raters – in such instances, the most frequently awarded score was recorded. These ratings have a Fleiss' Kappa score of 0.98 which indicates inter-rater reliability was exceedingly strong.

Since the number of participants being scored varied from Year Group to Year Group (depending upon how many had correctly identified the target



punchline), a mean score was calculated for each Year Group's verbal explanations. '2' was the highest possible score that could be obtained in the verbal explanation task. Mean explanation scores are illustrated in Figure 2.



*Figure 2. Mean explanation scores by year group*

As Figure 2 shows, the overall mean scores for verbal explanations increased as the Year Groups ascended: 0.92 (Year 2), 1.49 (Year 4), 1.63 (Year 6). An analysis of variance was run to determine whether significant differences existed between overall mean scores. This proved to be the case: ( $F(2, 57) = 14.16, p < 0.01$ ). The difference between Years 2 and 4 (0.57) was greater than that between Years 4 and 6 (0.14), however. Hence analyses of variance were run on Year 2 and 4 explanation scores, and Year 4 and 6 explanation scores, to determine whether differences between consecutive Year Groups were statistically significant. The difference (here an ascending increase) was statistically significant between Years 2 and 4 ( $F(1, 57) = 26.86, p < 0.01$ ) but not between Years 4 and 6 ( $F(1, 57) = 1.45, p > 0.05$ ). This mirrored the finding from the multiple-choice task in which an accelerated rate of development in the ability to identify ambiguity-based punchlines was established as occurring between Years 2 and 4.

As shown in Table 3, each ambiguity type was more frequently correctly identified as Year Groups ascended except for lexical ambiguity. Year 4 identified this type of ambiguity more frequently than Year 2 (as was the overall trend) but also more than Year 6 (which was out of keeping with the trend). The difference in the number of times lexical ambiguity was correctly identified by Year 4 over Year 6 was small (not significant), however, in relation to overall totals (55 times as opposed to 51) and did not affect the overall order in which ambiguity types were most frequently identified correctly.

Table 3

*Correct Punchline Selections by Ambiguity Type in the Multiple-Choice Task*

	<b>Lexical</b>	<b>Phonological</b>	<b>Morphological</b>	<b>Syntactic</b>	<b>Idiomatic</b>	<b>Standard deviation per Year Group</b>
<b>Year 2</b>	37 (62%)	24 (40%)	28 (47%)	22 (37%)	36 (60%)	6.12
<b>Year 4</b>	55 (92%)	53 (88%)	46 (77%)	38 (63%)	49 (82%)	5.98
<b>Year 6</b>	51 (85%)	54 (90%)	53 (88%)	49 (82%)	58 (97%)	3.03
<b>Totals</b>	143 (79%)	131 (73%)	127 (71%)	109 (61%)	143 (79%)	12.55

Lexical and idiomatic ambiguities were identified most frequently overall (although see later discussion on misleading identification rates for idiomatic ambiguity), followed by phonological, morphological, and syntactic ambiguities, respectively. A two-way analysis of variance was run on Year Groups and ambiguity types to determine whether ambiguity type had a significant effect on correctly identified original punchlines. Although the analysis of variance showed there to be no statistically significant effect for ambiguity type upon correct identification rates across the three Year Groups ( $F(4,8) = 3.52, p > 0.05$ ) the raw data nonetheless suggested some type of interaction. An analysis of variance was therefore run on each participating Year Group: Year 2 ( $F(4,76) = 26.03, p < 0.01$ ), Year 4 ( $F(4,76) = 6.05, p < 0.01$ ), Year 6 ( $F(4,76) = 2.21, p > 0.05$ ). There was a statistically significant relationship between scores and ambiguity type for Years 2 and 4, but not for Year 6. Ambiguity type thus affected comprehension significantly (in terms of identification rates) for the two youngest Year Groups only. These two Year Groups both identified lexical ambiguity correctly most frequently and syntactic ambiguity least frequently. There was variation in identification rates for the three remaining ambiguity types, but when scores were combined for Years 2 and 4, lexical ambiguities were identified most frequently (92 times), followed by idiomatic (85), phonological (77) morphological (74) and syntactic (60) ambiguities, respectively. This mirrored the overall findings for the multiple-choice task in that lexical and idiomatic ambiguities were identified jointly most frequently overall, and first and second most frequently respectively when scores for Years 2 and 4 were combined.



Explanations were scored 0, 1 or 2 depending on whether raters judged one, two (or no) parts of the ambiguity to have been explained. The maximum potential score for an explanation was 2. Mean scores (Table 4) were calculated for each Year Group’s scored explanations for each ambiguity type tested and accommodated the fact that only correct selections were scored by raters.

Table 4  
*Mean scores by ambiguity type in the verbal explanation task*

	<b>Ambiguity Type</b>					
	<b>Lexical Ambiguity</b>	<b>Phonological Ambiguity</b>	<b>Morphological Ambiguity</b>	<b>Syntactic Ambiguity</b>	<b>Idiomatic Ambiguity</b>	<b>All Ambiguity Types combined</b>
<b>Year 2</b>	1.00	1.04	0.84	0.78	0.89	0.92 (SD = 0.097)
<b>Year 4</b>	1.65	1.60	1.35	1.49	1.31	1.49 (SD = 0.134)
<b>Year 6</b>	1.70	1.75	1.62	1.53	1.54	1.63 (SD = 0.087)

As with overall explanation scores, mean scores for each ambiguity type increased with ascending Year Group and therefore age. The largest jump in mean scores for each ambiguity type occurred between Year 2 and Year 4. A two-way analysis of variance was run to determine whether differences in mean scores between Year Groups and between ambiguity types were significant. The analysis of variance showed differences to be statistically significant between both Year Groups ( $F(4,638) = 3.78, p < 0.05$ ) and ambiguity types ( $F(4,638) = 3.78, p < 0.05$ ). Of the five types of ambiguity tested, phonological ambiguities were found to score highest overall, followed by lexical, morphological, syntactic, and idiomatic ambiguities, respectively. That phonological and lexical ambiguities were comprehended more than morphological and syntactic ambiguities paralleled findings from the multiple-choice task. There was, however, a notable discrepancy in findings for the two tasks which related solely to the comprehension of idiomatic ambiguity. Data from the multiple-choice task showed that a high number of participants could correctly identify original punchlines based on idiomatic ambiguity but the subsequent analysis of scored explanations from the verbal explanation task contradicted this finding. More specifically, analysis of the transcriptions revealed that participants gave almost twice as many literal meanings of idiomatic phrases as figurative meanings (105 compared

to 56). This suggests that many of the correctly identified idiomatic punchlines had been chosen based on literal meanings only and not for the intended wordplay contained within punchlines. (It is worth noting that when original punchlines were rejected, participants tended to opt instead for a plausible punchline, as often accommodated in the literal interpretation of idiomatic words or phrases). Using multiple choice together with verbal explanations afforded a wider insight into idiomatic ambiguity comprehension than a single method would have achieved. Using mixed methods in this way meant that potential limitations regarding idiomatic ambiguity comprehension were addressed whilst improving upon previous modes of data collection for this ambiguity type.

Accommodating the above findings into the final analysis, results from both the multiple choice and verbal explanation task corresponded as to the types of ambiguity types comprehended most readily by participants. Lexical and phonological ambiguities were comprehended with most facility (although in opposing order in the two tasks) followed by morphological, syntactic, and idiomatic ambiguities, respectively. Notably, ambiguity types that required additional levels of processing (morphological, syntactic, and idiomatic) were comprehended less successfully than those that did not (lexical and phonological ambiguities).

#### **4. Discussion**

The current study aimed (a) to establish how children's comprehension of orally delivered verbal riddles differs across the ages of six to eleven years; (b) to determine whether ambiguity type affects the facility with which verbal riddles are understood by these children and (c) to establish whether our data can be explained within McGhee's 2002 model of humour development, or whether modification can be suggested. These points are now discussed in turn.

##### *4.1. Development of comprehension of verbal riddles according to school year*

Results from both the multiple choice and verbal explanation tasks showed that the ability to identify and explain ambiguities increased with age but that this increase was not equally spread across Year Groups. Differences in ambiguity identification and explanation scores were far greater between Years 2 (aged six to seven) and 4 (aged eight to nine) than between Years 4 (aged eight to nine) and 6 (aged ten to eleven) - and statistically significant only between Years 2 and 4. Thus, whilst there is an incremental increase in the ability to identify and explain ambiguities across all participating Year Groups, it was statistically significant only for those aged six to nine years. The results therefore support our first hypothesis that comprehension of ambiguity types will improve across ascending Year Groups. Results additionally showed an accelerated rate of improvement to occur between Years 2 (6-7) and Years 4 (8-9) that does not occur between Years 4 (8-9) and Year 6 (10-11).





#### 4.2. *Development of comprehension of verbal riddles according to ambiguity type*

Not all ambiguities were comprehended with the same facility and the facility with which they were comprehended varied according to the different ways in which linguistic properties were manipulated. This finding supports our second hypothesis - that the different ways in which linguistic properties are manipulated to create humour will affect participants' understanding, as we now evaluate.

Lexical ambiguities were identified most frequently overall in the multiple-choice task and syntactic ambiguities the least frequently (although see earlier discussion regarding misleading identification scores for idiomatic ambiguity). The finding that lexical ambiguities were identified most frequently overall and syntactic ambiguities least frequently was duplicated when scores were combined for Year Groups 2 and 4, both of whom had the greatest range in scores across ambiguity types and for whom findings were significant. Lexical ambiguity, in the current context, lay within individual lexical items and contained no grammatical class violations. This finding corresponds with those of Shultz (1974) and Hirsh-Pasek et al. (1978) both of whom, by virtue of the examples they provide, interpret it in the same way as the current study and report it as scoring highest in explanation tasks. Consistency in interpretation and application of what constitutes a lexical ambiguity therefore shows a similar trend in findings when compared across these studies.

After lexical ambiguity, the next highest scoring ambiguity in the verbal explanation task was phonological ambiguity – and this ambiguity type also comprised the second most identified type of ambiguity. This corresponds with the findings of Shultz and Pilon (1973) and Shultz and Horibe (1974) who also report phonological and lexical ambiguities as scoring more highly than other types of ambiguities – it must be borne in mind, however, that they include in the category of phonological ambiguity both homophones and confusion about word boundaries (Shultz & Pilon 1973, p.728, Shultz & Horibe 1974, p.14). Such linguistic phenomena are treated in the present investigation as comprising lexical and morphological ambiguities respectively.

Ambiguities which required processing above lexical (lexeme) and phonological (phoneme substitution) levels – i.e., morphological, syntactic, and idiomatic ambiguities - all required additional levels of processing and were all more difficult for participants to identify.

Morphological ambiguities requiring the (re)processing of word structure(s) and boundaries in addition to semantic knowledge relating to individual lexemes and the processing of sounds, were more difficult than lexical and phonological ambiguities for participants to identify. This parallels the findings of Hirsh-Pasek et al. (1978) who interpret 'morpheme boundary ambiguity' in a similar fashion and who also report it as being more difficult to explain than phonological or lexical ambiguities.

In ambiguities where grammatical class violations also occurred (e.g., a noun became a verb) participants were required to focus on the grammatical relationships between words at phrase level as well as at lexeme level to

understand the humour, resulting in this type of ambiguity being one of the hardest to process. It was least frequently identified and scored second lowest in the explanation task. This parallels the findings of Shultz and Pilon (1973) and Shultz and Horibe (1974), both of whom report 'deep structure' ambiguities as being difficult for young participant to explain. It also duplicates the findings of Yuill (1998), who reports syntactic ambiguities as being more difficult for children to identify<sup>5</sup>. It should be noted, however that not all Yuill's syntactic ambiguities are included in the class of syntactic ambiguity; she classifies 'Why do leopards make rubbish thieves? Because they're always spotted' as lexical ambiguity rather than syntactical (see earlier discussion), which means she tests this type of linguistic phenomena in more than one category within her study.

That syntactic ambiguity was least frequently identified (other than idiomatic ambiguity) and hardest to explain, does not correspond with Hirsh Pasek et al.'s (1978) findings for 'case-labelling'<sup>6</sup> ambiguities which they report as being easiest for children to explain (after lexical ambiguities). The examples they provide of this ambiguity type, however, show that some of their 'underlying structure' ambiguities are those which would be classified as either lexical or idiomatic ambiguities in the current study. They do not involve a change in word class and include fixed phrases which rely upon both figurative and literal interpretation to elicit humour.

Idiomatic ambiguity, which involved the processing of figurative meanings in addition to literal ones, was the most difficult ambiguity type for participants to comprehend, although this finding was not always reflected in the multiple-choice scores. We note that the reason for this is that the punchlines for this ambiguity type were often wrongly chosen based on their literal meanings only. The bias towards literal meanings of idiomatic expressions supports findings from previous studies on children's idiom comprehension - Ackerman (1982), Prinz (1983), Gibbs (1987), Levorato & Cacciari (1995), Le Sourn-Bissaoui et al. (2012) all report that children have a bias towards literal interpretations of figurative language when processing idioms. This bias is influenced by immature cognitive and language skills (Levorato & Cacciari, 1995) which weaken around seven years of age when children start to acquire new skills in decoding figurative expressions.

Given the late and protracted development of non-literal processing skill, it is unsurprising that idiomatically ambiguous original punchline selections were made solely based on their literal meanings by so many of the younger children in the current study who were not cognitively advanced enough to be able to fully process figurative language. Of note is the fact that each of the idiomatic ambiguities contained within riddles was low in transparency. This meant that it was harder for participants to decompose the idioms when trying to make sense of them. The meanings of the idioms 'a piece of cake',

---

<sup>5</sup> Of the seven types of riddle punchlines Yuill opts to test, not all specifically rely on verbal ambiguities, however. She also tests 'pragmatic riddles', 'absurdity riddles' and 'metalinguistic riddles'. Only three of Yuill's classifications match those used in the current study and some labels differ (Yuill's word-compound' ambiguity corresponds to our 'morphological' ambiguity).

<sup>6</sup> Hirsh-Pasek's 'case-labelling and underlying structure' relates to differences in word class as incorporated by our syntactic ambiguity category.



‘going up the wall’ and ‘having a screw loose’ could not be determined from their constituent parts alone and required previous exposure in order they be understood. It is possible that low transparency may have contributed to them being harder for participants to process than those higher in transparency. This would correspond with the findings of Gibbs (1987), Nippold and Rudzinski (1993), Levorato and Cacciari (1999), Nippold and Taylor (2002) and Cain, Oakhill and Lemmon (2005). All the idiomatic ambiguities included in the study had been trialled in the pilot study, however, to ensure that participants of this age were familiar with their meanings – which they were. That they were low in transparency is therefore less likely to have been a factor in our findings than the fact that idiomatic ambiguities (of any type) involve more complex processing skills than those required to process other ambiguity types and as such are a later language skill to be acquired by children, which is reflected in the outcomes. Comparing our findings for idiomatic ambiguity with earlier studies is difficult since this ambiguity type is not reported as having been overtly tested. Lack of reported findings do not mean that this type of ambiguity has not been tested previously, however. It has, although researchers have failed to address the fact that it requires a specific type of competence and have instead included idiomatic ambiguities in differing categories (Fowles & Glanz in lexical ambiguity and Shultz & Pilon in phonological ambiguity). We have, in our study, ensured this specific linguistic phenomenon – and requisite processing skills – has been robustly accommodated within a category of its own and this has allowed us to determine the impact that this type of linguistic manipulation has on children’s developing ambiguity and humour comprehension.

As our discussion shows, there are both parallels and discrepancies to be drawn between current outcomes and reported findings from previous studies on children’s humour and ambiguity comprehension. Ambiguity types have not always been interpreted in the same way, however, which means that whilst parallels may be drawn, they do not always correspond to the type of linguistic phenomena tested. Discrepancies also occur where parallels might otherwise have been drawn had ambiguities been interpreted in the same way. This highlights the need for consistency in ambiguity classifications to facilitate comparative analysis across studies which would then enrich our understanding of the way(s) in which different language phenomena affect children’s developing ability to comprehend ambiguities. This issue was addressed here by using definitions which focused specially upon inherent language phenomena embedded within the form(s) of verbal riddles.

Outcomes from the two tasks show that the decoding of various ambiguity types requires different processing strategies, and that these, in turn, depend upon the type of language phenomena exploited to elicit humour. Some ambiguity types are more complex to understand because they require additional levels of processing. Current findings show that the less complex the processing demand, the more readily the ambiguity is understood (phonological and lexical ambiguities). The more complex the processing demand, the less readily it is understood (morphological, syntactic, and idiomatic ambiguities respectively). This supports our third hypothesis: that

ambiguities requiring more complex language processing skills will be less readily understood.

Our findings will now be assessed against McGhee's current model of humour development and revisions will be made accordingly.

#### 4.3. *McGhee's Model of humour development*

Although comprehension results for Year 2 were lower than for Years 4 and 6, there were, nonetheless, a considerable number of Year 2 children (aged 6-7) who could identify and explain the ambiguity in some of the riddles. This finding therefore corresponds with McGhee's current framework of humour development in which children, typically around seven years of age are said to make the transition from Stage Four to Stage Five humour and to begin to develop the ability to comprehend ambiguity-based jokes.

Whilst our findings support McGhee's model of humour development as just described, it nonetheless highlights a specific aspect of the current framework now open to challenge. Participants in Year 2, despite having lower identification rates than the two older Year Groups, were shown to be capable of identifying almost half of the original punchlines correctly (147 times out of a potential 300 = 49%). Even when potentially misleading identification rates for idiomatic ambiguity (see Table 2) were removed, participants in Year 2 were still able to correctly identify 111 punchlines. Participants in this Year Group ranged from six to seven years. Given that they were able to correctly identify almost half the punchlines containing ambiguous words/phrases, these participants are well on their way to being able to identify ambiguity-based punchlines as opposed to simply "begin[ning]" (McGhee, 1979, p.76) to develop understanding. Even allowing for individual variation, this proficiency is earlier than McGhee's model would suggest. Therefore, whilst the current data confirms the biggest developmental leap in understanding verbal ambiguities to occur between the ages of six and nine years, it also suggests that children might start consolidating their understanding of ambiguity-based humour at an earlier age than previously purported. Whether this was because the participants were cognitively advanced, linguistically advanced, or because the link between cognitive and humour development is not as closely bound as previously reported is not ascertainable from the current study but certainly merits further investigation. Future studies, for example, might build on this finding by testing participants younger than those that participated here to determine whether the age at which young children start to comprehend ambiguities is earlier than presently accounted for in McGhee's five-stage framework of humour development. Should this be the case, then McGhee's model (first formulated some forty years ago) might be modified accordingly.

#### 4.4. *Developing the humour model*

In addition to both supporting and challenging the humour model, outcomes from the multiple choice and verbal explanation tasks provided an insight into trends in verbal ambiguity comprehension not currently accommodated or communicated by the model in its present form. As participants made the transition to Stage Five humour, they found some types of ambiguities easier/more difficult to comprehend. All ambiguity types required



participants to assign two different meanings to a single word/string of words in order that they be understood, but the processes involved in retrieving and assigning the two meanings varied according to the linguistic phenomena upon which ambiguity types were based. This, in turn, had a bearing on the types of ambiguity that were identified and explained most successfully. Given this finding, it is therefore proposed (Table 5) that a set of sequential sub-stages are added to McGhee's Stage Five humour model. Each proposed sub-stage (5a: lexical & phonological ambiguities; 5b. morphological ambiguities; 5c. syntactic ambiguities & 5d. idiomatic ambiguities) interrelates with both the language phenomena embedded within the wording of verbal riddle (i.e., the form in which it is presented) and the different processing demands it makes of the listener.

Table 5  
*Revised stages of humour development*

---

Stage 0 (Laughter without humour: 0 to 6 months) – Stage 4 (Conceptual incongruity: 3 to 5 years)
Stage 5: Riddles and jokes (double meanings/ambiguities: 6-7 to 10-11 Years)
<i>Stage 5a:</i> lexical and phonological ambiguities
<i>Stage 5b:</i> morphological ambiguities
<i>Stage 5c:</i> syntactic ambiguities
<i>Stage 5d:</i> idiomatic ambiguities

---

It is noted that of these four sub-stages (5a-5d), the final three stages comprise one type of ambiguity only, whilst the first stage comprises two: lexical and phonological ambiguities. Both were understood more than the three other types (morphological, syntactic, and idiomatic) in the multiple choice and explanation tasks - but in an opposing order. It therefore remains to be determined as to whether one type is more readily understood than the other as children first start to comprehend ambiguity-based humour. This might thus prove an area for future focused investigation to refine the humour model further.

## **5. Conclusion**

Our study leads us to conclude that the facility with which verbal ambiguities are comprehended by young children varies according to the different ways in which linguistic phenomena have been manipulated to elicit humour. We have accordingly applied our findings to Stage 5 of McGhee's humour model within which verbal ambiguities first start to be comprehended, to accommodate children's developmental progress within this specific stage.

## **Acknowledgement**

This work was supported by the Economic and Social Research Council (grant number ES/J500197/1)

**References**

- Ackerman, B. (1982). On Comprehending Idioms: Do Children Get the Picture? *Journal of Experimental Child Psychology*, 33, 439-454. [https://doi.org/10.1016/0022-0965\(82\)90058-3](https://doi.org/10.1016/0022-0965(82)90058-3)
- Allington, R., & Strange, M. (1979). Remembering is Not Necessarily Understanding in Content Areas. *Reading Horizons* 20 (1), 60-64. [https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=2125&context=reading\\_horizons](https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=2125&context=reading_horizons)
- Attardo, S., & Raskin, V. (1991). Script theory revis(it)ed: joke similarity and joke representation model. *Humor: International Journal of Humor Research* 4 (3-4), 293- 348. <https://doi.org/10.1515/humr.1991.4.3-4.293>
- Baker, G., & Aldridge, M. (2021). Disambiguating ambiguity: Providing a Framework for Classifying Types of Ambiguity. *Linguistics and the Human Sciences* 14 (3), 237-260. <https://journal.equinoxpub.com/LHS/article/view/19339>
- Bariaud, F. (1989). Age Differences in Children's Humor. *Journal of Children in Contemporary Society* 21 (112), 15-45. [https://doi.org/10.1300/J274v20n01\\_03](https://doi.org/10.1300/J274v20n01_03)
- Berger, K. (2011). *The Developing Person Through the Life Span (eighth edition)*. New York: Worth Publishers.
- Cain, K., Towse, A., & Knight, R. (2009). The Development of Idiom Comprehension: An Investigation of Semantic and Contextual Processing Skills. *Journal of Experimental Child Psychology*, 102, 280-298. <https://doi.org/10.1016/j.jecp.2008.08.001>
- Cain, K., Oakhill, J., & Lemmon, K. (2005). The Relation between Children's Reading Comprehension Level and Their Comprehension of Idioms. *Journal of Experimental Child Psychology*, 90, 65-87. <https://doi.org/10.1016/j.jecp.2004.09.003>
- Fowles, B., & Glanz, M. (1977). Competence and Talent in Verbal Riddle Comprehension. *Journal of Child Language*, 4 (3), 433-452. <https://doi.org/10.1017/S0305000900001781>
- Gibbs, R., Jr. (1987). Linguistic Factors in Children's Understanding of Idioms. *Journal of Child Language*, 14, 569-586. <https://doi.org/10.1017/S0305000900010291>
- Graham, E. E., Papa, M. J., & Brooks, G. P. (1992). Functions of humor in conversation: Conceptualization and measurement. *Western Journal of Communication*, 56 (2), 161-183. <https://doi.org/10.1080/10570319209374409>
- Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Hirsh-Pasek, K., Gleitman, L., & Gleitman, H. (1978). What did the Brain Say to the Mind? A Study of the Detection and Report of Ambiguity by Young Children. In A. Sinclair, R Jarvella & W. Levelt (Eds.) *The Child's Conception of Language* (pp.97-132). SpringerVerlag. [https://doi.org/10.1007/978-3-642-67155-5\\_7](https://doi.org/10.1007/978-3-642-67155-5_7)
- Hoff, E. (2015). Language Development. In M. Bornstein & M. Lamb (Eds.), *Developmental Science (Seventh edition)* (pp. 443-488). New York and London: Psychology Press.



- Hoppe, R., Kess, J. F. (1982). *The Acquisition of Metalinguistic Abilities*. Paper presented at the International Conference of the AILA Commission on Psycholinguistics, "First and Second Language Learning: Similarities and Difference" (Milan, Italy).
- Hyter, Y. D. (2017). Pragmatic assessment and intervention in children. In L. Cummings (Ed.), *Research in clinical pragmatics* (pp. 493–526). Cham: Springer. [https://link.springer.com/chapter/10.1007/978-3-319-47489-2\\_19](https://link.springer.com/chapter/10.1007/978-3-319-47489-2_19)
- Kaplan, P. (1998). *The Human Odyssey*. St Paul, MN: West Publishing Company
- Le Sourn-Bissaoui, S., Caillies, S., Bernard, S., Deleau, M., & Brulé, L. (2012). Children's Understanding of Ambiguous Idioms and Conversational Perspective-Taking. *Journal of Experimental Child Psychology*, 112, 437-451.  
[https://www.researchgate.net/publication/224865848\\_Children's\\_understanding\\_of\\_ambiguous\\_idioms\\_and\\_conversational\\_perspective-taking](https://www.researchgate.net/publication/224865848_Children's_understanding_of_ambiguous_idioms_and_conversational_perspective-taking)
- Levorato, C., & Cacciari, C. (1995). The Effects of Different Tasks on the Comprehension and Production of Idioms in Children. *Journal of Experimental Child Psychology*, 60, 261-283.  
<https://doi.org/10.1006/jecp.1995.1041>
- Levorato, M. & Cacciari, C. (1999). Idiom Comprehension in Children: Are the effects of semantic analysability and context separable. *Journal of Cognitive Psychology*, 11 (1), 55-66.  
<https://doi.org/10.1080/713752299>
- McGhee, P. (1971a). The Role of Operational Thinking in Children's Comprehension and Appreciation of Humor. *Child Development*, 42 (3), 733-744. <https://doi.org/10.2307/1127444>
- McGhee, P. (1971b). Cognitive Development and Children's Comprehension of Humor. *Child Development*, 42 (1), 123-138.  
<https://www.jstor.org/stable/1127069>
- McGhee, P. (1972). On the Cognitive Origins of Incongruity Humor: Fantasy Assimilation Versus Reality Assimilation. In P. McGhee & J. Goldstein (Eds.), *The Psychology of Humor* (pp. 61-80). New York: Academic Press
- McGhee, P. (1977). A Model of the Origins and Early Development of Incongruity-Based Humor. In A. Chapman & H. Foot (Eds.), *It's a Funny Thing Humor* (pp. 27-36). Oxford: Pergamon Press
- McGhee, P. (1979). *Humor, its Origin and Development*. San Francisco: W. H. Freeman and Company
- McGhee, P. (2002). *Understanding and Promoting the Development of Children's Humor*. Iowa: Kendall/Hunt Publishing Company
- McGhee, P., & Panoutsopoulou, T. (1990). The Role of Cognitive Factors in Children's Metaphor and Humor Comprehension. *Humor: International Journal of Humor Research*, 3 (4), 379-402.  
<https://doi.org/10.1515/humr.1990.3.4.379>
- Nezlek, J. B., & Derks, P. (2001). Use of humor as a coping mechanism, psychological adjustment, and social interaction. *Humor: International Journal of Humor Research*, 14(4), 395–413.  
<https://doi.org/10.1515/humr.2001.011>

- Nippold, M. & Rudzinski, M. (1993). Familiarity and Transparency in Idiom Explanation: A Developmental Study of Children and Adolescents. *Journal of speech and hearing Research*, 36 (4), 728-737. <https://doi.org/10.1044/jshr.3604.728>
- Nippold, M. A., & Taylor, C. L. (2002). Judgments of idiom familiarity and transparency: A comparison of children and adolescents. *Journal of Speech, Language, and Hearing Research*, 45(2), 384-391. [https://doi.org/10.1044/1092-4388\(2002/030\)](https://doi.org/10.1044/1092-4388(2002/030))
- Prentice, N., & Fathom, R. (1975). Joking Riddles: A Developmental Index of Children's Humor. *Developmental Psychology*, 11 (2), 210-216. <https://doi.org/10.1037/h0076455>
- Prinz, P. (1983). The Development of Idiomatic Meaning in Children. *Language and Speech*, 26 (3), 263-272. <https://doi.org/10.1177%2F002383098302600307>
- Raskin, V. (1985). *Semantic Mechanisms of Humor*. Dordrecht-Boston-Lancaster: D. Reidel.
- Shultz, T. (1974). Development of the Appreciation of Riddles. *Child Development*, 45, 100-105. <https://www.jstor.org/stable/1127755>
- Shultz, T., & Horibe, F. (1974). Development of the Appreciation of Verbal Jokes. *Developmental Psychology*, 10 (1), 13-20. <https://doi.org/10.1037/h0035549>
- Shultz, T. & Pilon, R. (1973). Development of the Ability to Detect Linguistic Ambiguity. *Child Development*, 44, 728-733. <https://doi.org/10.2307/1127716>
- Suls, J. (1972) A Two-Stage Model for the Appreciation of Jokes and Cartoons: An Information-Processing Analysis. In Goldstein, J. and McGhee, P. (Eds). *The Psychology of Humor. Theoretical Perspectives and Empirical Issues*, (pp. 81-100). New York: Academic Press
- Suls, J. (1983). Cognitive Processes in Humor Appreciation. In McGhee, P. & Goldstein, J. (Eds.) *Handbook of Humor research Volume 1: Basic Issues*, (pp. 39-57). New York: Springer-Verlag.
- Tunmer, W.E., Pratt, C. & Herriman, M. L. (Eds.) (1984) *Metalinguistic Awareness in Children: Theory, Research, and Implications*. Springer Series in Language and Communication, Vol. 15. Berlin: Springer-Verlag.
- Wolfstein, M. (1954). *Children's Humor: A Psychological Analysis*. Bloomington: Indiana University Press.
- Yalisove, D. (1978). The Effect of Riddles Structure on Children's Comprehension of Riddles. *Developmental Psychology*, 14 (2), 173-180. <https://doi.org/10.1037/0012-1649.14.2.173>
- Yuill, N. (1998). Reading and Riddling: The Role of Riddle Appreciation in Understanding and Improving Poor Text Comprehension in Children. *Current Psychology of Cognition*, 17 (2), 313-342. <http://sro.sussex.ac.uk/id/eprint/13114>.
- Zipke, M. (2007). The Role of Metalinguistic Awareness in the Reading Comprehension of Sixth and Seventh Graders. *Reading Psychology*, 28 (4), 375-396. <https://doi.org/10.1080/02702710701260615>





- Zipke, M. (2008). Teaching Metalinguistic Awareness and Reading Comprehension with Riddles. *The Reading Teacher*, 62 (2), 128-137. <https://core.ac.uk/download/pdf/56325349.pdf>
- Zipke, M. (2009). Using Semantic Ambiguity Instruction to Improve Third Graders' Metalinguistic Awareness and Reading Comprehension: An Experimental Study. *Reading Research Quarterly*, 44 (3), 300-321. <https://www.jstor.org/stable/25655457>
- Ziv, A. (1976). Facilitating Effects of Humor on Creativity. *Journal of Educational Psychology*, 68 (3), 318-322. <https://doi.org/10.1037/0022-0663.68.3.318>
- Ziv, A. (1983). The Influence of Humorous Atmosphere on Divergent Thinking. *Contemporary Educational Psychology* 8, 68-75. [https://doi.org/10.1016/0361-476X\(83\)90035-8](https://doi.org/10.1016/0361-476X(83)90035-8)
- Ziv, A. (1988). Using Humor to Develop Creative Thinking. *Journal of Children in Contemporary Society*, 1-2, 99-116. [https://doi.org/10.1300/J274v20n01\\_07](https://doi.org/10.1300/J274v20n01_07)

## Appendices

### Appendix A: Ambiguity definitions

#### **Lexical Ambiguity**

Lexical ambiguity occurs solely within the alternative meaning of an individual lexical item and does not rely upon grammatical analysis at phrase, clause or sentence level. It occurs when an individual word has more than one meaning without any class violation. This type of ambiguity encompasses both homonyms and polysemes since when relayed orally, both carry the same sound but different meanings e.g. "Why are babies good at football?" "Because they can **dribble**."

#### **Phonological Ambiguity**

Phonological ambiguity occurs when the ambiguous fragment of riddle text has two non-identical phonetic forms for the two alternative interpretations. The modification of the phonetic form can comprise the addition, deletion, or substitution of a phoneme. It does not involve modification of phonetic form across word boundaries and is contained within a single lexical item e.g. "What do whales eat for dinner?" "Fish and **ships**."

#### **Morphological Ambiguity**

Morphological ambiguity occurs when there are changes in morpheme boundaries for the two readings of the text. Other than variation in stress or juncture, the ambiguous fragment of the riddle has identical phonetic forms for the two alternative interpretations e.g. "Why did the jelly wobble?" "Because it saw the **milkshake/milk shake**."

#### **Syntactic Ambiguity**

Syntactic ambiguity occurs when two different underlying syntactic structures are mapped onto a single surface structure. The two different syntactic representations reflect different underlying grammatical relations between lexical items. Syntactic ambiguity relies upon grammatical analysis at whole phrase, clause, or sentence level e.g. "How was the blind carpenter able to see?" "He picked up his hammer and **saw**."

#### **Idiomatic Ambiguity**

Idiomatic ambiguity occurs when the figurative meaning of an idiom is blended with the literal meanings of its individual lexical components e.g. "What does Spiderman do when he's angry?" "He goes **up the wall**."



## Appendix B: Riddle Questions with Multiple Choice Punchline Selections

**Key**

Original punchline

**Plausible punchline**

*Irrelevant punchline*

*Riddle 1 (Idiomatic Ambiguity)*

**Why did the robot act silly?**

**Because he was in a daft mood**

Because he had a screw loose

*Because he liked apples*

*Riddle 2 (Syntactic Ambiguity)*

**Why is six afraid of seven**

*Because seven eats fish*

Because seven ate/eight nine

**Because seven is bigger**

*Riddle 3 (Morphological Ambiguity)*

**Why couldn't the skeleton go to the ball?**

He had no body/nobody to go with

*He was too cold*

**It was past his bedtime**

*Riddle 4 (Lexical Ambiguity)*

**Why are babies good at football?**

Because they can dribble

**Because they kick their legs**

*Because they like music*

*Riddle 5 (Phonological Ambiguity)*

**How did the banana know he was ill?**

He wasn't peeling well

**He had a high temperature**

*He looked out of the window*

*Riddle 6 (Idiomatic Ambiguity)*

**What does spiderman do when he's angry?**

*He turns on the radio*

He goes up the wall

**He stamps his feet**

Riddle 7 (Morphological Ambiguity)

**Why did the jelly wobble?**

**Because someone shook the plate**

Because it saw the milk shake/milkshake

*Because it was midnight*

Riddle 8 (Syntactic Ambiguity)

**How was the blind carpenter able to see?**

*He went to the circus*

He picked up his hammer and saw

**He put on his glasses**

Riddle 9 (Lexical Ambiguity)

**When is the best time to buy chickens?**

*When the tide is out*

**When they are fresh**

When they're going cheap/cheep

Riddle 10 (Phonological Ambiguity)

**What do whales eat for dinner?**

Fish and ships

**Tasty sea creatures**

*Big earrings*

Riddle 11 (Syntactic Ambiguity)

**Why do leopards make rubbish thieves?**

*Because they smell of roses*

**Because they always get caught**

Because they're always spotted

Riddle 12 (Idiomatic Ambiguity)

**Why did the schoolboy eat his homework?'**

**His friend said it tasted nice**

His teacher said it was a piece of cake

*His mum liked singing in the bath*

Riddle 13 (Lexical Ambiguity)

**Why can't you ever win at cards in the jungle?**

**Because there are so many wild animals competing**

*Because it snows on the mountains*

Because there are too many cheetahs/cheaters

Riddle 14 (Morphological Ambiguity)

**When are roads angry?**

*When the birds are singing*

When they are cross roads/crossroads

**When you annoy them**



*Riddle 15 (Phonological Ambiguity)*

***What's a mouse's favourite game?***

Hide and squeak

*Time for bed*

**Hunt the Cheese**