OPEN ACCESS

UNIVERSITY of BRISTOL

Peer reviewed version

University of Bristol - Explore Bristol Research

General rights

Preschoolers Joke with Jokers, but Correct Foreigners

Elena Hoicka[1,2] & Nameera Akhtar[1]

[1]University of California, Santa Cruz, [2]University of Stirling

Correspondence concerning this article should be addressed to Elena Hoicka,

Psychology Department, University of Stirling, Stirling, FK9 4LA, Scotland, U.K. E-

mail: elena.hoicka@stir.ac.uk Tel: +44 (0) 1786467649 Fax: +44 (0) 1786 467641.

Abstract

Thirty- and 36-month-old English speakers' ($N = 106$) ability to produce jokes, distinguish between humorous and sincere intentions, and distinguish between English- and Foreign-language speakers, was examined in two tasks. In the Giving task, an experimenter requested one of two familiar objects, and a confederate always gave her the wrong object. In the Naming task, the confederate mislabeled familiar objects. In the English-speaking conditions, the confederate laughed after doing the wrong thing (English-Humor) or said, "There!" (English-Sincere). In the Foreign conditions, the French or Italian-speaking confederate laughed (Foreign-Humor) or said, "D'accord!" or "Va bene!" (Foreign-Sincere). When preschoolers were subsequently requested to give and name the same objects and a new set of familiar objects they were significantly more likely to imitate and "do the wrong thing" in the Humor versus Sincere, and in the English versus Foreign conditions.

Keywords: Intentions, Humor, Imitation, Foreign languages, Trust, Creativity

Preschoolers Joke with Jokers, but Correct Foreigners

Children understand intentions from a young age. Infants complete others'
incomplete actions from 15 months (Meltzoff, 1995). From 14 months, they
discriminate between incidental and intentional aspects of an action (Gergely,
Bekkering, & Király, 2002). Infants can also differentiate intentional actions from
mistakes from 14 months (Carpenter, Akhtar, & Tomasello, 1998). Most research on
intention understanding compares whether children can tell the difference between
intentional and unintentional actions. However very little research has examined
whether young children understand different types of intentions. For example, people
can intend to be sincere, or they can intend to joke. People can also be mistaken
because they do not speak or understand the language being used. The current study
examines whether preschoolers can differentiate between humorous and sincere
intentions, and between native and foreign speakers.

By 2 years, toddlers begin to understand not only that people can intend to do
the right thing, but also that people can intend to do the wrong thing. Hoicka and
Gattis (2008) found that from 2 years toddlers will copy wrong actions accompanied
by laughter, but correct the same wrong actions accompanied by the expression
"Whoops!" Rakoczy, Tomasello, and Striano (2004) found that 36-month-olds will
copy technically wrong pretend actions marked with laughter amongst other cues, but
will correct the same wrong actions when marked by trying-but-failing cues (e.g.,
grunting).

However, intentional actions were cued with positive expressions in studies
examining intentional versus unintentional actions (e.g., "There!" in Carpenter, et al.,
1998; laughter in Hoicka & Gattis, 2008; Rakoczy et al., 2004), while unintentional
actions were cued by negative expressions ("Whoops!" in Carpenter, et al., 1998;

Hoicka & Gattis, 2008; grunting in Rakoczy et al., 2004). Thus it is possible that infants and toddlers responded to emotion cues rather than the underlying intentions. Therefore children may have copied actions associated with positive emotions and avoided or corrected actions associated with negative emotions. Consequently, it is important to examine whether children can distinguish between intentions that are accompanied by similar emotion cues.

Recent research suggests that 15-month-old infants can distinguish between two different types of positive intentional acts. Infants were shown sweet actions (e.g., stroking a toy cat) as well as humorous actions (e.g., stroking one's head with a toy cat) accompanied by matching or mismatching vocal cues (Hoicka & Wang, 2010). The infants looked longer when the cues mismatched the actions than when they matched, indicating that they were sensitive to the mismatch. Thus specific cues (in particular, laughter and humorous intonation) led infants to expect humor, while equally positive sweet cues (saying "Awww" and sweet intonation) led infants to expect sweet (but non-humorous) actions. The findings suggest that 15-month-olds distinguish between humorous and sweet cues, and match them appropriately to humorous and sweet actions.

One goal of the current study was to examine whether preschoolers distinguish between different intentions that all involve positive emotion. Intentional actions typically involve doing the right thing. However humorous intentional actions often involve doing the wrong thing (Hoicka & Gattis, 2008; Hoicka, Jutsum, & Gattis, 2008; Leekam, 1991). As such, humor provides an ideal way to compare various intentions that involve positive emotions. In the case of labeling, one could intentionally mislabel an object because one is making a joke. Thus the joker would have intended to do the wrong thing, and would have fulfilled that intention.

A second goal was to examine whether children can discriminate between native and foreign speakers, and understand that it is incorrect to use a foreign language in an English-speaking context. In contrast to joking, someone might mislabel objects in an English-speaking context because she genuinely does not know the correct label in English. Thus while the (mis)labeling is an intentional act, it results in doing the wrong thing. There is evidence that young children can distinguish between native and foreign speakers. For example, 6-month-olds preferred to look at people who spoke their own language versus another language, and 5-year-olds preferred to befriend people who spoke their own language versus another language (Kinzler, Dupoux, & Spelke, 2007). Additionally, 3- and 4-year-olds are more likely to learn novel words from an English speaker versus a foreign-language speaker (Akhtar, Menjivar, Hoicka, & Sabbagh, 2010). Furthermore, preschoolers selectively imitate the actions of native versus foreign-accented speakers (Kinzler, Corriveau, & Harris, in press). One reason that children might be less likely to copy the behaviors of a foreigner is that they may not trust them, as they could be seen as generally inaccurate speakers. Indeed, 3- and 4-year-olds are less likely to trust inaccurate versus accurate speakers when learning new labels (Koenig & Harris, 2005).

A third goal was to determine whether toddlers can be innovative and produce their own jokes. Past research suggests that young children do produce jokes. For example, Hoicka and Gattis (2008) found that from 19 months, children will copy incongruous actions (e.g., putting a boot on one's hand) accompanied by laughter. Observational studies have described several types of humor that individual children have produced, such as mislabeling objects or making up silly words (e.g., Hoicka & Akhtar, 2010; Johnson & Mervis, 1997), or performing incongruous actions such as bending over and putting one's head through one's legs, or brushing a pot (e.g.,

Hoicka & Akhtar, 2010; Loizou, 2005). Parents report that from 8 months infants will repeat incongruous actions when others laugh (Reddy, 2001). However, it is possible that children in these studies simply mimic or repeat particular actions and utterances in order to get a response (e.g., laughter). This would not necessarily require children to understand that what makes a joke funny is incongruity (e.g., McGhee, 1979; Shultz, 1974), that is, something unexpected or out of the ordinary. In order to determine whether young children appreciate incongruity as humorous, and to determine whether young children can create their own jokes, experimental methods are required. To date, there have been no experimental studies to test whether young children can create novel jokes. In the current studies, we tested not only whether children would copy specific joke tokens, but also whether children would create their own novel joke tokens.

In the current study, we compared children's responses to adults who all deliberately "did the wrong thing" (gave the wrong object, or mislabeled an object) in a positive manner. There are different reasons why an adult might do the wrong thing. One could be that they are joking (cued by laughter versus a sincere expression). Another could be that they do not speak the language (cued by speaking a foreign language versus English). We employed a 2 (Humorous, Sincere) X 2 (Foreign, English) design to examine whether children would be more likely to do the wrong thing (give the wrong object, say the wrong label) in the presence of cues that the speaker/actor was a foreigner and/or was intending to be humorous.

It is possible that children in past studies (e.g., Carpenter, et al., 1998; Hoicka & Gattis, 2008; Rakoczy, et al., 2004) responded to emotional valence only, without understanding underlying intentions. Thus we might expect children in the current study to do the wrong thing equally often across conditions, as all conditions in the

current study are cued with positive emotion. However children in past studies may have indeed been sensitive to intentions, not just emotion. Thus we might expect children in the current study to be sensitive to humorous intentions and understand that it is funny to purposefully do the wrong thing. They may also understand that it makes no sense to do the wrong thing in the sincere conditions. If this is the case, children should do the wrong thing in the two Humor conditions, but not in the two Sincere conditions.

Likewise, if children respond to the confederate's language, they should be more likely to correct the confederate in the Foreign versus English conditions. This is because foreign-language speakers violate English-language conventions, and may appear to be ignorant and hence not trustworthy. Overall, we might expect children to do the wrong thing most often in the English-Humor condition, least often in the Foreign-Sincere condition, and somewhere in between in the English-Sincere and Foreign-Humor conditions.

In the Giving task, 30- and 36-month-olds watched an experimenter ask a confederate for one (e.g., car) of two known objects (e.g., car, book). The confederate always gave the incorrect object to the experimenter (in this case, the book). In the Naming task, children watched the experimenter ask the confederate to label familiar objects (e.g., duck). The confederate always mislabeled the object with a silly (made-up) word (e.g., *Boody doody*). After having a chance to copy or correct the confederate's actions/ utterances, children were asked to give/label a new set of familiar objects. This allowed us to examine whether children would continue to do the wrong thing in the Humor versus Sincere conditions, demonstrating that (1) children were not simply mimicking the confederate, but instead understood that she was intending to be humorous, and (2) they were able to create their own novel jokes.

We tested 30- and 36-month-olds because 2-year-olds are theorized to enjoy jokes based on mislabeling (McGhee, 1979), and 3-year-olds are theorized to make jokes that play with concepts (McGhee, 1979) including nonsense language, e.g., saying, "gobbledygook" (Johnson & Mervis, 1997). Indeed, Hoicka and Akhtar (2010) found, through parent interviews and observation of parent-child dyads, that 2- and 3-year-olds do produce such jokes to some extent.

Method

Participants

One hundred and twelve children participated. Six children were not included because they did not speak English (2), did not respond on any trial (2) or responded in only one block of trials throughout the entire study (2). Of the remaining 106 children, 27 were assigned to the English-Humor condition (15 boys; 13 were 30 months, $M = 29$ months, 25 days, $SD = 20$ days; 14 were 36 months, $M = 36$ months, 1 day, $SD = 23$ days), 27 to the English-Sincere condition (12 boys; 12 were 30 months, $M = 29$ months, 26 days, $SD = 19$ days; 15 were 36 months, $M = 35$ months, 24 days, $SD = 20$ days), 25 to the Foreign-Humor condition (10 boys; 12 were 30 months, $M = 29$ months, 28 days, $SD = 22$ days; 13 were 36 months, $M = 36$ months, 4 days, $SD = 20$ days), and 27 to the Foreign-Sincere condition (12 boys; 13 were 30 months, $M = 29$ months, 28 days, $SD = 20$ days; 14 were 36 months, $M = 36$ months, 4 days, $SD = 23$ days). Children were tested in California or Scotland. One child was Asian/Pacific Islander, 89 children were Caucasian, two were Hispanic, 10 were of mixed ethnicity, and four parents did not report their child's ethnicity. Children were recruited from a press release in a local newspaper, music and gymnastic classes, children's fairs, and libraries. Two of the 106 children did not participate in the Giving task. Twenty-five of the 106 children did not participate in the Naming task. A

mixed logit model (explained later) looking at whether or not children participated in the Naming task was not affected by Intention, Language, Age, Gender, Task Order, or the number of task object words produced, as reported by parents.

Materials

Materials for the Giving task included 12 pairs of familiar objects, e.g., car, book; pizza, toothbrush (see Appendix A for full list). Materials for the Naming task included 12 different familiar objects, e.g., duck, cup, elephant, hairbrush (see Appendix B for full list). Objects were chosen for which 90% of 30-month-olds should know the labels according to the MacArthur-Bates Child Development Inventories (MB-CDI, Dale & Fenson, 1996). Parents reported that the 104 children who participated in the Giving task understood on average 23.89 out of 24 words (99.54%). Parents reported that the 81 children who participated in the Naming task could produce 11.49 out of the 12 words on average (95.75%). A Sony digital DV camcorder, and a wall-mounted microphone were used to record the sessions.

Design

The study consisted of a 2 (Language: English, Foreign) X 2 (Intention: Sincere, Humorous) between-subjects design. The confederate spoke English in the warm-up in the English conditions, and French or Italian in the Foreign conditions. The confederate laughed after giving the wrong object (Giving) or saying the wrong label (Naming) in the Humorous conditions, and said either "There!" (English-Sincere) or "D'accord!" or "Va bene!" (Foreign-Sincere) while smiling in the Sincere conditions. The dependent measures were whether children gave the correct or incorrect object on each trial (Giving), and whether children labeled objects correctly or incorrectly on each trial (Naming). Order of task (Giving, Naming) was

counterbalanced such that half the children in each condition received the Giving task first.

For the Giving task, two sets of six pairs of familiar objects were counterbalanced such that half of the children saw the confederate model objects from Set 1, while the other half saw the confederate model objects from Set 2 (see Appendix A). The experimenter asked the confederate, and later child, for the first object of each pair (e.g., asked for the car from the car/book pair) for half the children, and asked for the second object of each pair (e.g., asked for the book from the car/book pair) for the other half. Placement of objects (left or right) was counterbalanced within and across children and across conditions. Pairs of objects were used in a fixed order (see Appendix A).

For the Naming task, two sets of six individual familiar objects were counterbalanced such that half of the children saw the confederate label objects from Set 1, while the other half saw the confederate label objects from Set 2 (see Appendix B).

Procedure

To warm up, the experimenter, confederate, and child engaged in a 5- to 10-minute free play session until the child was comfortable interacting with the experimenter and confederate. In the English conditions, the experimenter and confederate both spoke English during the warm-up, but neither the experimenter nor the confederate labeled any objects, referring to them as "that one" or "him", and so on. In the Foreign conditions, the confederate spoke French or Italian, and the experimenter spoke English but did not label any objects. While children did not generally make verbal requests, the experimenter and confederate avoided responding to verbal requests and instead tried to follow the children's physical actions. For

example, if a child put an animal in a barn, the experimenter or confederate did as well. While some children in the Foreign conditions preferred to play with the experimenter, perhaps avoiding the confederate because of the language she spoke, other children interacted well with the confederate, seemingly not noticing the language barrier.

One of four scripts was used by the experimenter and confederate before beginning the session in order to communicate that they were making jokes or playing a game (see Appendix C). For example, in the Humor-English condition, the experimenter (E) and confederate (C) said the following:

E: You know, [confederate] is really funny.

C: I love to be silly!

E: How about we make some jokes! I'll ask you for some things and you two can be silly.

Giving Task

The experimenter placed a pair of objects (e.g., a car and a book) in front of the confederate and asked, "Do you want to give me the [object label] (e.g., car)?" This wording was used instead of, "Can you give me the [object label]?" as it permitted the children to give the wrong object if desired when it was their turn. The confederate always gave the incorrect object (in this case, the book). In the Humor conditions, the confederate then laughed and looked to the child's face, followed by the experimenter and the child's parent also laughing and looking to the child's face. In the Sincere conditions, the confederate said "D'accord!" or "Va bene!" (Foreign-Sincere) or "There!" (English-Sincere), smiled, and looked to the child's face, after which the experimenter said "Alright!" smiled and looked to the child's face, and the

parent said, "OK!" smiled and looked to the child's face. The same pattern followed for an additional five pairs of objects.

Identical trials. Each child received six trials with the same objects the confederate had seen. The experimenter set out the pairs of objects in the same order for the child (one pair at a time), and requested the same objects as requested from the confederate (although the left/right placement of objects was varied). When the child gave an object, the experimenter said. "OK" followed by a slight laugh. If the experimenter had not laughed when the child attempted a joke, the absence of laughter could have served as feedback, stopping the child from joking. However the feedback needed to be consistent regardless of condition or response, so the experimenter gave a slight laugh in all conditions and for all responses. This could be interpreted as humor appreciation or general jolliness, making the response appropriate either for a joking or a general play event. If the child did not give an object, the experimenter requested the object again. If the child still did not give an object, the experimenter went on to the next pair of objects.

Extension trials. The experimenter then brought out six additional pairs of familiar objects that had not been seen previously. The experimenter continued to request one of two objects, and gave the same feedback as previously.

Naming Task

The experimenter held up an object and asked the confederate, "What do you want to call this?" The confederate always answered with a made-up label, in all conditions (e.g., Boody doody; Oogle boo, see Appendix B). The confederate, experimenter and parent then all gave the same cues by condition as in the Giving task. This was repeated for an additional five objects.

Identical trials. The experimenter then set out the same objects for the child (one at a time), and asked the child, "What do you want to call this?" This wording was used instead of, "What do you call this?" or "What is this called?" as it permitted the child to produce an incorrect label if desired. When the child responded, the experimenter responded as in the Giving task by saying "OK" followed by a slight laugh, or repeating the question if the child did not respond.

Extension trials. The experimenter then brought out a new set of six familiar objects for the extension trials. The experimenter continued to ask the child, "What do you want to call this?" and gave the same feedback as previously.

Coding

For the Giving task, actions were coded from video by coders blind to the hypotheses as (1) correct (child gave the requested object); (2) incorrect (child gave the non-requested object); or (3) non-response (gives neither, or both at same time). For example, if the experimenter set out a car and a book and requested the car, the child's response was coded as correct if she gave the car, and incorrect if she gave the book. If the child gave one object, then another (whether immediately or later), the first object was coded. Twenty-two of the 104 videos (21.15%) were coded for agreement. Cohen's kappa was excellent, $k = .94$.

For the Naming task, labels for each object were transcribed from video. Labels were coded as (1) correct (child produced the correct label, e.g., "duck" for duck); (2) incorrect (child produced the incorrect label, e.g., "*Boody doody*", or "cup" for duck; see Appendix D for details of coding scheme); or (3) no response (child said nothing, "I don't know"; "Yeah"; "No"; etc.). Twenty-two of the 89 videos (24.72%) were coded for agreement. Cohen's kappa was very good, $k = .89$. A qualitative analysis examined the types of incorrect labels used by the children in the English-

Humor condition. This involved classifying incorrect labels as nonsense language (e.g., "mooka" instead of "spoon"), wrong words (e.g., "dog" instead of "spoon"), or a play on the actual word (e.g., "spoony spoon spoon" instead of "spoon").

In order to examine whether children appreciated that the confederate was joking in the Humor conditions, we also examined whether children laughed at some point during each trial. We expected them to laugh more often in the Humor conditions than in the Sincere conditions, and more often when they gave incorrect objects (Giving) and produced incorrect labels (Naming) than when they gave the requested objects and labeled correctly. We coded whether or not children laughed (1) each time the confederate did something wrong, and (2) each time the child responded during the identical and extension trials. Eighteen of the 106 videos (16.98%) were coded for agreement. Cohen's kappa was good, $k = .780$. For disagreements on all coding, we followed the coding of the primary coder.

Results

We modeled the likelihood of making incorrect responses using logit mixed effects models with the LME4 package (Bates, Maechler, & Dai, 2008) in R (R Development Core Team, 2009). It has been the convention to use ANOVA for collapsed data across trials, and ANOVA has been used in similar studies (e.g., Hoicka & Gattis, 2008; Meltzoff, 1995; Rakoczy, et al., 2004). However Jaeger (2008) has recently argued that it is more appropriate to analyze this type of data with logit mixed effects models, as the children's responses are repeated-measures categorical data points, not continuous data points. Thus, just as one would use a chi-square analysis instead of a t-test when the variable is categorical rather than continuous, the same can be done for repeated-measures variables. Jaeger (2008) demonstrates that ANOVAs on this type of data can lead to spurious results through

over-fitting, while logit mixed effects models avoid this. Additionally logit mixed effects models use all raw data instead of collapsed data, giving the models more power. ANOVAs on each child's percentage of incorrect responses can be found in Appendix E. Comparing this to the following mixed logit models, it can be seen that most of the main results are the same. However the mixed logit models also capture an effect of Language in the Giving task, perhaps revealed through taking account of other variables not measurable in the ANOVA such as Trial Number, as well as the increased power offered by using raw data instead of collapsed data.

For each analysis (Giving, Naming, Laughter), we first built a base model, which included an intercept, and Participant and Items (1-12 from both sets of items for each of the Giving and Naming tasks) as random variables. We compared the base model to models including Intention (Humor, Sincere), Language (English, Foreign), Trial Type (Identical, Extension), Age, Gender, Task Order (Giving or Naming task first) and Trial Number (earlier versus later trials) all with $df = 1$. If one of these models was a significant improvement over the base model (using chi-square analyses), the variable was added to the base model. Otherwise the base model was not changed. If more than one variable was a significant improvement over the base model, the model which had the lowest log-likelihood was used as the model, and we then checked whether the remaining variables improved this new model in the same way. We also tested whether interactions of variables with each condition, and response in the laughter analysis, significantly improved the latest model in the same way. Once there were no improvements to the model where $df = 1$, we looked for improvements where $df = 2$. For example, if Intention (Humor, Sincere) was in the model, but Gender was not, then looking at an interaction between Intention and Gender would be $df = 2$ (Gender; Gender X Intention). When the best $df = 2$

interaction was found, we again tested whether variables and interactions with $df = 1$ improved the model, and repeated the cycle. If interactions with $df = 2$ did not improve the model, then we examined whether interactions with $df = 3$ improved the model. For example, if neither Intention nor Gender were in the best model, examining an interaction between Intention and Gender would have $df = 3$ (Intention; Gender; Intention X Gender). When the best $df = 3$ interaction was found, we again tested whether variables and interactions with $df = 1$ improved the model, and repeated the cycle. If interactions with $df = 3$ did not improve the model, then the best-fit model had been found. When the final best-fit model was found, the logit mixed effects model determined which variables and interactions had significant effects on the children's responses, and returned an odds-ratios (OR) which describes whether certain variables and interactions make the target response proportionally more or less likely.

Giving Task

See the left half of Figure 1 for the percentage of trials on which children gave incorrect objects in the Giving task, by Condition. The base model was improved by adding Intention (Humor, Sincere), $X^2(1) = 4.51$, $p = .0338$, and an interaction of Language (English, Foreign) and Trial Order, $X^2(3) = 9.87$, $p = .0197$, as fixed effects. The resulting model (*log-likelihood* = -538.05, $N = 1148$) included significant effects of Intention (Humor vs. Sincere, $OR = 2.35$, $p = .0288$), Language (English vs. Foreign, $OR = 179.59$, $p = .0236$), Trial Order ($OR = 1.20$, $p = .0053$), and an interaction between Language and Trial Order ($OR = 1.27$, $p = .0115$). Thus children in the Humor conditions were more than twice as likely to give an incorrect object as children in the Sincere conditions. Children in the Foreign conditions were 180 times as likely to give the correct object as children in the English conditions. Children were

16

20 percent more likely to give the wrong object during later versus earlier trials. Finally, children were 27 percent more likely to give the correct versus incorrect object on later trials during the Foreign versus English conditions. No other effects were found.

Naming Task

See the right half of Figure 1 for the percentage of trials on which children produced incorrect labels in the Naming task, by Condition. The base model was improved by adding Trial Type (Identical, Extension), $X^2(1) = 6.09$, $p = .0136$, then Order (Giving or Naming first), $X^2(1) = 6.01$, $p = .0142$, then Language (English, Foreign), $X^2(1) = 4.79$, $p = .0286$, then Intention (Humor, Sincere), $X^2(1) = 4.05$, $p = .0441$ as fixed effects. The resulting model (*log-likelihood* = -341.5, $N = 813$) included significant effects of Intention (Humor vs. Sincere, $OR = 3.70$, $p = .0383$), Language (English vs. Foreign, $OR = 4.00$, $p = .0314$), Trial Type (Identical vs. Extension, $OR = 1.76$, $p = .0106$), and Order (Giving or Naming first, $OR = 0.22$, $p = .0161$). Thus children in the Humor conditions were almost 4 times as likely to label objects incorrectly as children in the Sincere conditions. Children in the English conditions were 4 times as likely to label objects incorrectly as children in the Foreign conditions. Children were almost twice as likely to label objects incorrectly during the Identical versus Extension Trials. Finally, children were about a fifth as likely to label objects incorrectly when they did the Giving task before the Naming task. No other effects were found.

Types of Incorrect Labels in the English-Humor Condition

From a total of 110 incorrect labels generated in the English-Humor condition, none were exactly the same as those produced by the confederate. Seventy-eight (71%) were in the form of nonsense words, for example, a duck was called, "*colgai*",

a diaper was called, "*wooga*", and a cup was called, "*goojooboojoo*". Twenty-one (19%) of the incorrect labels involved calling objects the incorrect name, e.g., a watch was called "diaper", and a brush was called "spider monster". Finally, 11 (10%) of the incorrect labels involved playing with the actual label by repeating the label, adding nonsense syllables or irrelevant words, or rhyming the real labels with silly labels. For example, a block was called, "*blocky block*", while a baby doll was called, "*baby daby*".

Laughter

Secondary analyses examined (1) whether children laughed more often when the confederate did the wrong thing in the Humor versus Sincere conditions, which reveals whether children were processing the confederate's behavior as joking, and (2) whether children laughed more often when doing the wrong thing versus doing the right thing, which reveals whether children were appreciating and/or cuing their own jokes. Since laughter was a somewhat rare event, we collapsed data across the Giving and Naming tasks for these analyses. One child's data could not be coded for laughter as her face could not be seen.

Laughter: Modeled trials.

See Figure 2 for the percentage of Modeling trials during which children laughed, by Intention (Humor, Sincere) and Age. A logit mixed effects model was built as before, however there were 24 items included in the random effects rather than 12 since both the Giving and Naming tasks were included. Additionally, we investigated fixed effects of Task (Giving, Naming) and its interaction with Condition. The dependent measure was whether or not children laughed. The base model was improved by adding Intention (Humor, Sincere), $X^2(1) = 22.70$, $p < .0001$, then Trial Number, $X^2(1) = 14.33$, $p = .0002$, then Task, $X^2(1) = 9.43$, $p = .0021$, then

Age, $X^2(1) = 6.16$, $p = .0131$, and finally an interaction between Intention (Humor, Sincere) and Age, $X^2(1) = 4.40$, $p = .0360$ as fixed effects. The resulting model (*log likelihood* = -284.46, *N* = 1248) included significant effects of Trial Number (*OR* = 1.42, $p < .0001$), and Task (Giving, Naming, *OR* = 2.19, $p = .0019$), and a trend for an interaction between Intention (Humor, Sincere) and Age (*OR* = 1.71, $p = .0931$). Thus children were more than twice as likely to laugh in the Naming task than in the Giving task. Children were almost fifty percent more likely to laugh on later versus earlier trials. Finally, a trend suggests that older children were more than seventy percent more likely to laugh in the Humor versus Sincere conditions as younger children. No other effects were found.

Laughter by response type.

See Figure 3 for the percentage of Identical and Extension trials during which children laughed, by Response and Age. A logit mixed effects model was built as for the Modeling trials, however we included children's Response Type (Correct, Incorrect) as a variable, as well as Trial Type (Identical, Extension). The base model was improved by adding Response Type, $X^2(1) = 7.22$, $p = .0072$, then an interaction between Response Type and Age, $X^2(2) = 16.51$, $p = .0003$, then an interaction between Response Type and Task (Giving, Naming), $X^2(2) = 11.51$, $p = .0032$, then an interaction between Intention (Humor, Sincere) and Task, $X^2(2) = 13.63$, $p = .0011$, then an interaction between Intention (Humor, Sincere) and Trial Type, $X^2(2) = 6.32$, $p = .0425$ and finally an interaction between Intention (Humor, Sincere) and Response Type, $X^2(1) = 11.75$, $p = .0006$ as fixed effects. The resulting model (*log likelihood* = -181.59, *N* = 1937) included a significant interaction between Response Type and Age (*OR* = 1.77, $p = .0006$). Thus older children were almost twice as likely

to laugh when giving an incorrect versus correct response than younger children. No other reliable effects were found.

Discussion

The results suggest that children are sensitive to (1) humorous versus sincere intentions, and (2) English versus foreign languages. Additionally, they suggest that children can create their own novel jokes. During the Giving task, children were significantly more likely to give the wrong object in the humorous conditions, even during the extension trials, in which the confederate had not modeled the actions. Likewise, during the Naming task, children were significantly more likely to produce an incorrect label in the humorous conditions, even during the extension trials. Children were significantly less likely to give an incorrect object and to produce an incorrect label when the confederate spoke a foreign language.

Understanding Different Types of Intentions

One important aspect of the design of this study was that all contexts were emotionally positive, but involved different intentions - intentions to do the wrong thing (Humor) versus intentions to do the right thing (Sincere). In past studies (e.g., Carpenter, et al., 1998; Hoicka & Gattis, 2008; Rakoczy, et al., 2004), all intentional actions, including pretense and humor, involved positive emotional displays, while mistakes and trying involved negative emotional displays. Thus, one possibility is that the children in these studies were not in fact discerning others' intentions. Rather, they may have copied behaviors when the demonstrator was positive, and avoided or corrected behaviors when the demonstrator was negative. Another possibility is that children really did respond to the underlying intentions (e.g., as in Hoicka & Wang, 2010), as the intentional cues in these studies matched the action types modeled. The

current findings suggest that preschoolers respond to underlying intentions rather than emotions only, and can also differentiate between different types of intentions.

A limitation of the current study is that laughter is more intense than the expressions, "There!" "D'accord!" and "Va bene!" accompanied by smiling. Thus children might copy actions marked with high intensity positive emotions, and avoid those marked with low intensity. However expressions such as "There!" encourage copying in other studies (e.g., Carpenter, et al., 1998). Thus children might follow a set of behavioral rules: (1) for novel acts with low intensity positive emotion, copy; (2) for wrong acts with low intensity positive emotion, correct; (3) for novel acts with high intensity positive emotion, copy. Future research should compare positive intentional cues matched for intensity to rule out this possibility.

Another question remains as to whether laughter alone cues humor, or whether the social setting is also important. In the humor conditions, the experimenter, confederate, and parent all laughed. While children as young as 19 months will copy wrong actions when the experimenter, and the experimenter alone, laughs (Hoicka & Gattis, 2008), other studies suggest that the presence of a laughing audience plays a role in children's appreciation of humor. For example, 7- and 8-year-olds judge materials as funnier when other people are present (Chapman, 1975). Young children may also better trust that the intention was to joke when parents were willing to laugh. Infants are more likely to laugh when mothers, versus strangers, produce jokes involving putting on masks, and peek-a-boo (MacDonald & Silverman, 1978; Waters, Matas, & Sroufe, 1975). Future research should explore the social factors that influence children's ability to infer humorous intentions.

Children in the Foreign-language conditions were more likely to correct the confederate in the Giving and Naming tasks. Children may have thought that the

Foreign-language speaker intended to say the right thing, but did not know how. Thus children may have corrected the Foreign-language speaking confederate because they were trying to match her intention – to do the right thing. In contrast, children may have thought that the English-speaking confederate should know the correct labels, and hence had intended to say the wrong thing. Thus children may have been more likely to mislabel during the English conditions because children might have tried to match their intentions to that of the confederate – to do the wrong thing.

A further consideration is that children may have thought the confederate in the Sincere conditions was lying rather than being sincere. Thus it is possible that children did not differentiate intentions to do the right thing from intentions to do the wrong thing, but instead differentiated two types of intentions to do the wrong thing. This should theoretically be more difficult to do, and past research suggests that children cannot differentiate jokes from lies, albeit in a much more verbal form, until they understand false belief, at around 4 to 5 years (Leekam, 1991). Future research should examine whether preschoolers differentiate intentions to joke, lie, and be sincere.

Trust

Another way to interpret these results is through the framework of trust as young children often do not trust people who do the wrong thing (Koenig & Harris, 2005), prefer people who speak their own language over people who speak a different language (Kinzler, et al., 2007), and selectively imitate the actions of native over foreign-accented speakers (Kinzler, et al., in press). Again, this may reflect the fact that children do not trust information coming from someone who says the wrong thing, or says it in the wrong way. In the context of speaking English, speaking French or Italian might appear to be wrong. However, the current findings suggest

that there is one context in which children will trust others who do the wrong thing; in a humorous context, children understand that incorrect actions can sometimes be performed intentionally (Hoicka & Gattis, 2008).

We suggest that children in the Humor conditions were more likely to copy and extend incorrect actions and labels because they understood that in a humorous context (i.e., when a joker is intentionally doing the wrong thing), it is appropriate to do the wrong thing themselves. However in the Sincere conditions, since the confederate gave cues that she intended to do the right thing, children may have seen her as an inaccurate labeler, and hence did not trust, and did not copy her wrong actions. Finally, in the Foreign conditions, children may have thought that the foreigner was an inaccurate labeler, and hence should not be trusted, nor copied. Thus, a key to whether children are willing to trust a wrong-doer may be whether or not the wrong-doer intended to do the wrong thing. If someone gives cues that they intended to do the right thing, but did the wrong thing instead, children may see the action either as a mistake or based in ignorance, as children are aware that people do the wrong thing for these reasons (e.g., Hoicka & Gattis, 2008; Koenig & Harris, 2005). Thus intentional cues on their own may not be enough for a child to copy an action. Rather, it may be the combination of intentional cues and whether the intentional cues match that person's actions or utterances which influence a child's willingness to do the wrong thing.

Creating Novel Jokes

Children differentiated conditions in the same way for both identical and extension trials in both the Giving and Naming tasks, suggesting they coordinated their actions with the context of the game, rather than the individual actions of the confederate. Thus preschoolers were not simply matching humorous intentional cues

to particular actions or utterances, but actually understood the nature of what made the actions humorous: that they involved giving the incorrect object, or producing an incorrect label. Thus, preschoolers could create joke tokens that they had never seen following a specific joke type they observed (give the wrong thing/say the wrong thing). Indeed, when children in the English-Humor condition generated incorrect labels for the objects, either in the Identical or Extension trials, they always made up novel incorrect labels for the objects. This reveals that they understood the incongruity inherent to the jokes, and also demonstrates that preschoolers can be creative when being humorous.

These are the first experimental studies to show that preschoolers can create their own jokes, demonstrating an ability to innovate. While much observational and interview-based research has suggested that toddlers and even infants can do this (e.g., Hoicka & Akhtar, 2010; Johnson & Mervis, 1997; Loizou, 2005; Reddy, 2001), such research could be biased by parents' or researchers' interpretations of infants' and toddlers' intentions as humorous. It is possible that young children were simply mimicking or repeating actions or utterances that appeared to be jokes because they got a positive response for doing so. But this does not necessitate that children understood that what they were doing was incongruous, and that incongruity is what makes things funny. The current research suggests that from at least 30 months preschoolers do attend to specific types of incongruity (giving the wrong object, uttering an incorrect label for an object) and use them to form their own jokes. As they had seen the confederate give incorrect objects and mislabel objects, it could be said that children were simply following a script to do similar things rather than being creative. However identifying that script involved identifying types of incongruities. Additionally, following the script in the Extension trials required children to apply the

script in ways they had never seen anyone else apply the script, which is a creative act. Indeed, in the case of mislabeling objects, the children created jokes that no one else has probably ever made, for example, calling a cup, "goojooboojoo". Thus from 30 months, preschoolers can create their own joke tokens based on joke types that they have seen modeled.

Laughter

Children were more likely to laugh during modeling trials on the Naming versus Giving task. This suggests children found the Naming task to be funnier. Additionally, children were more likely to laugh during later versus earlier modeling trials, suggesting that as children saw more examples of the same type of wrong action, children perhaps "got it" and found it funny. Older children were more likely to laugh when they did the wrong thing compared to younger children. This suggests that when older children performed a wrong act, they did so because they thought it was funny, and were either laughing at their own jokes, or cuing others to their jokes. Indeed, children laugh twice as much when joking versus not joking (Hoicka & Akhtar, 2010), and laugh more often when an act is an intentional joke, versus an incidental incongruity (Bainum, Lounsbury, & Pollio, 1984). This suggests a richer interpretation that children, at least the 36-month-olds, understood that they were joking, and not simply following the rules of the game.

<div align="center">

Conclusions

</div>

The current findings reveal that preschoolers understand that people can intend to do the wrong thing in order to be funny. In order to do this, they needed to distinguish intentions to do the wrong thing from intentions to do the right thing, all expressed along the same (positive) emotional spectrum. Thus preschoolers understand that people have different types of positive intentions. Additionally, we

found that preschoolers can create their own jokes, revealing their creativity, as well as a deeper understanding of what makes a joke funny. Finally, the findings also suggest that English-speaking preschoolers understand that when a foreign language speaker says or does the wrong thing, it is different from when a native English speaker does so. Thus children use a variety of cues and contexts in order to determine when it is appropriate to do the right thing, or do the wrong thing.

References

Akhtar, N., Menjivar, J., Hoicka, E., & Sabbagh, M. (2010). Learning novel labels from a foreign speaker: The role of experience with more than one language. *Submitted Manuscript.*

Bainum, C. K., Lounsbury, K.R., & Pollio, H. R. (1984) The development of laughing and smiling in nursery school children. *Child Development, 55,* 1946-1957.

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using S4 classes [Computer software manual]. Available from http://lme4.r-forge.r-project.org/ (R package version 0.999375-23)

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen through 18 month old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*, 315-330.

Chapman, A. (1975). Humorous laughter in children. *Journal of Personality and Social Psychology, 31,* 42-49

Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*, 125-127.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415,* 755.

Hoicka, E., & Akhtar, N. (2010). Spontaneous Humor Production in 2- and 3-year-olds. *Submitted Manuscript.*

Hoicka, E. & Gattis, M. (2008). Do the wrong thing: How toddlers tell a joke from a mistake. *Cognitive Development, 23,* 180-190.

Hoicka, E., Jutsum, S., & Gattis, M. (2008). Humor, abstraction, and disbelief. *Cognitive Science, 32,* 985-1002.

Hoicka, E., & Wang, S. (2010). Fifteen-month-olds match intentional cues to actions. *Submitted Manuscript.*

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434-446.

Johnson, K. E., & Mervis, C. B. (1997). First steps in the emergence of verbal humor: A case study. *Infant Behavior and Development, 20*, 187-196.

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (in press). Children's selective trust in native-accented speakers. *Developmental Science.*

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 12577-12580.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76,* 1261-1277.

Leekam, S. R. (1991). Jokes and lies: Children's understanding of intentional falsehood. In A. Whiten (Ed.). *Natural theories of mind: Evolution, development and simulation of everyday mindreading*. (pp. 159-174). Oxford: Basil Blackwell.

Loizou, E. (2005). Infant humor: The theory of the absurd and the empowerment theory. *International Journal of Early Years Education, 13,* 43-53.

MacDonald, N. E., & Silverman, I. W. (1978). Smiling and laughter in infants as a function of level of arousal and cognitive evaluation. *Developmental Psychology, 14,* 235-241.

McGhee, P. (1979). *Humor: Its origin and development*. San Francisco: W.H. Freeman and Company.

Meltzoff, A. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838-850.

Rakoczy, H., Tomasello, M., & Striano, T. (2004). Young children know that trying is not pretending: A test of the "behaving-as-if" construal of children's early concept of pretense. *Developmental Psychology*, *40*, 388–399.

R Development Core Team. (2009). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from http://www.R-project.org

Reddy, V. (2001). Infant clowns: The interpersonal creation of humor in infancy. *Enfance, 53,* 247-256.

Shultz, T. R. (1974). Development of the appreciation of riddles. *Child Development, 45*, 100–105.

Waters, E., Matas, L., & Sroufe, L. A. (1975). Infants' reactions to an approaching stranger: Description, validation, and functional significance of wariness. *Child Development, 46*, 348-356.

Figure 1. Percentage of all children's incorrect responses by Condition and Task.

Figure 2. Percentage of all modeling trials during which children laughed by Intention (Humor, Sincere) and Age.

Figure 3. Percentage of all identical and extension trials during which children laughed by Response Type (Incorrect, Correct) and Age.

Figure 1.

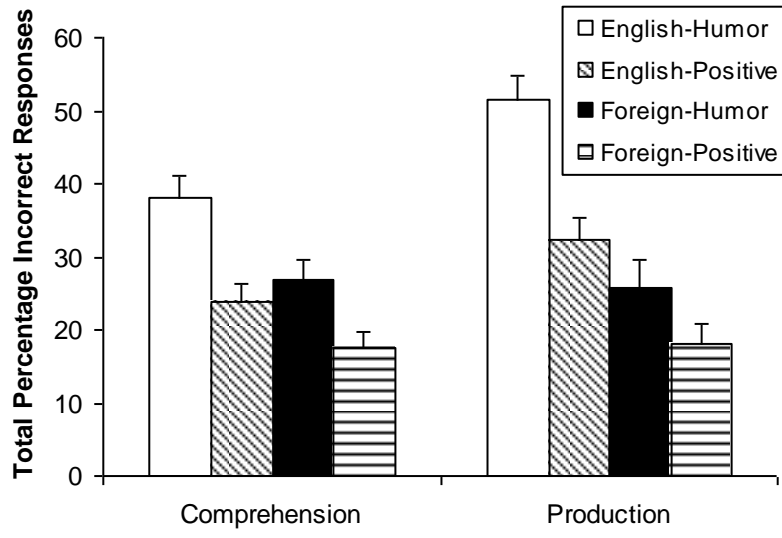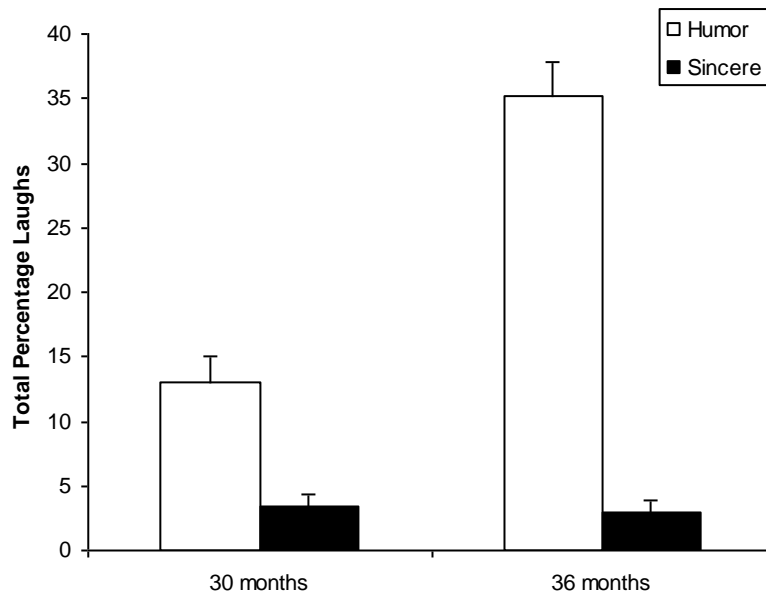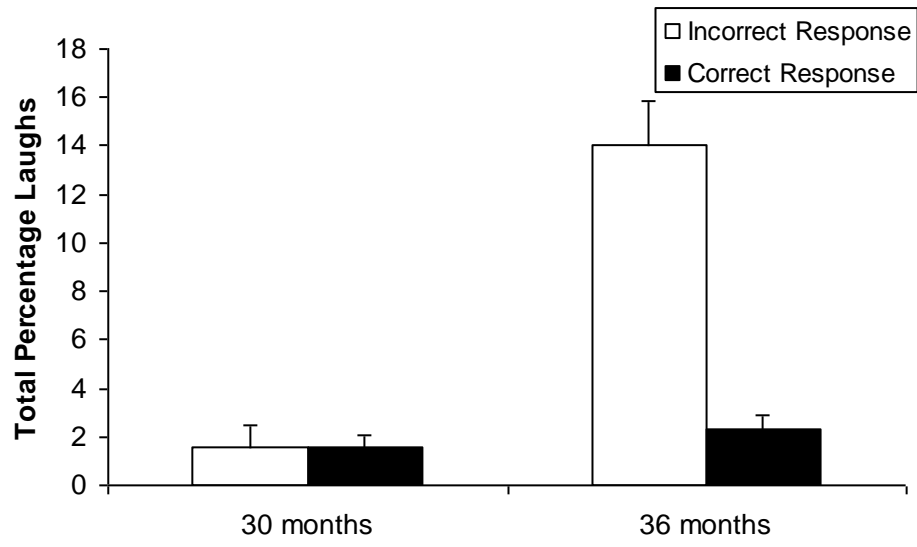Figure showing a bar chart. Y-axis labeled "Total Percentage Incorrect Responses" ranging from 0 to 60. X-axis has two groups: Comprehension and Production. Legend: English-Humor, English-Positive, Foreign-Humor, Foreign-Positive.

Figure 2.

Figure 3.

34

Appendix A
Object pairs used in Giving task.

Set 1:

	Set 1	Set 2
Pair 1	Book	Car
Pair 2	Boat	Ball
Pair 3	Pizza	Toothbrush
Pair 4	Bowl	Apple
Pair 5	Fish	Sock
Pair 6	Hat	Horse

Set 2:

	Set 1	Set 2
Pair 1	Bottle	Pencil
Pair 2	Balloon	Fork
Pair 3	Cat	Shirt
Pair 4	Pants (Trousers)	Dog
Pair 5	Bear	Keys
Pair 6	Plate	Pig

Appendix B

Objects used in Naming task, and the confederate's (incorrect) labels of the objects.

Set 1	Set 2	Confederate's Label
Elephant	Duck	Boody doody
Diaper (Nappy)	Shoe	Gaga head
Banana	Grapes	Oogle boo
Spoon	Cup	Moley schmoley
Doll	Block	Wuggy woo
Watch	Brush	Gobble gook

Appendix C

Scripts used to introduce communicative context of English-Humor, English-Sincere,

Foreign-Humor, and Foreign-Sincere conditions by Experimenter (E) and

Confederate (C).

Giving task

English-Humor:

    E: You know, [confederate] is really funny.

    C: I love to be silly!

    E: How about we make some jokes! I'll ask you for some things and you two

    can be silly.

English-Sincere:

    E: You know, [confederate] loves to have fun.

    C: I love to play games!

    E: How about we play a game. I'm going to ask you for some things and you

    two can help me.

Foreign-Humor:

    E: You know, [confederate] is really funny, and [confederate] doesn't speak

    much English

    C: J'aime bien faire des blagues! [French]

    E: How about we make some jokes! I'll ask you for some things and you two

    can be silly.

Foreign-Sincere:

> E: You know, [confederate] doesn't speak much English
>
> C: J'aime bien faire des blagues! [French]/ Va a discoteca! [Italian]
>
> E: How about we play a game. I'm going to ask you for some things and you
>
> two can help me.

Naming task:

English-Humor:

> E: You know, [confederate] is really funny.
>
> C: I love to be silly!
>
> E: How about we make some jokes! I'll ask you what some things are called
>
> and you two can be silly.

English-Sincere:

> E: You know, [confederate] loves to have fun.
>
> C: I love to play games!
>
> E: How about we play a game. I'm going to ask you what some things are
>
> called and you two can help me.

Foreign-Humor:

> E: You know, [confederate] is really funny, and [confederate] doesn't speak
>
> much English
>
> C: J'aime bien faire des blagues! [French]
>
> E: How about we make some jokes! I'll ask you what some things are called
>
> and you two can be silly.

Foreign-Sincere:

> E: You know, [confederate] doesn't speak much English
>
> C: J'aime bien faire des blagues! [French]/ Va a discoteca! [Italian]
>
> E: How about we play a game. I'm going to ask you what some things are
>
> called and you two can help me.

When either the Giving or Naming task came second, the first two lines were omitted, and the experimenter said the final portion of the script, except that she said, "How about we make some more jokes" or "How about we play another game".

Appendix D

Coding Scheme for the Naming Task

1. Correct

    a. The exact same word, e.g., calls elephant, "Elephant"

    b. Mispronounced correct word (drops syllable, mispronounces phoneme), e.g., calls elephant, "elphant", calls banana, "nana", calls grapes, "gwapes"

    c. A correct synonym or category, e.g., calls doll, "baby", calls diaper, "pull-up", calls elephant, "animal", calls banana, "fruit"

    d. Describes by correct color: e.g., calls elephant, "grey", calls banana, "yellow"

    e. Describes by correct noise, e.g., calls duck, "quack", calls watch, "tick-tock"

2. Incorrect

    a. Incorrect word, e.g., calls elephant, "monkey", calls spoon, "shirt"

    b. Says silly word, e.g., calls shoe, "moogy"

    c. Plays with real word by repeating it, rhyming it, adding extra sounds or words, e.g., calls cup "cup cup", calls brush "brushy goo", calls banana "lip stick banana"

Appendix E

Results for Giving and Naming Tasks using ANOVA on children's percentage of

incorrect responses.

Preliminary analyses found no effect of Gender or Age, so these were dropped

from the final analyses.

Giving

Preliminary analyses found no effect of Order, so it was dropped from the

final analysis. A 2 (Intention: Humorous, Sincere) X 2 (Language: English, Foreign)

between-subjects ANOVA found that children were significantly more likely to give

the incorrect object in the Humor ($M = 32.44\%$, $SD = 29.40\%$) than Sincere

conditions ($M = 20.79\%$, $SD = 26.12\%$), $F(1, 100) = 4.49$, $p = .037$, $\eta = .043$. There

was no difference between how often children gave the incorrect object in the English

($M = 29.25\%$, $SD = 30.47\%$) and Foreign conditions, ($M = 23.65\%$, $SD = 25.74\%$),

$F(1, 100) = 1.08$, $p = .302$.

Naming

A 2 (Intention: Humorous, Sincere) X 2 (Language: English, Foreign) X 2

Order (Comprension task first, second) between-subjects ANOVA found that children

were significantly more likely to produce incorrect labels in the Humor ($M = 41.94\%$,

$SD = 38.88\%$) than Sincere conditions ($M = 21.88\%$, $SD = 28.95\%$), $F(1, 73) = 4.76$,

$p = .032$, $\eta = .061$. Children in the English conditions ($M = 38.71\%$, $SD = 36.43\%$)

were significantly more likely to produce an incorrect label than children in the

Foreign Conditions ($M = 21.05\%$, $SD = 31.16\%$), $F(1, 73) = 5.87$, $p = .018$, $\eta = .074$.

Children were more likely to produce an incorrect label if they did the Naming task

second ($M = 39.09\%$, $SD = 35.35\%$) instead of first ($M = 22.90\%$, $SD = 33.52\%$), F(1,

73) = 4.51, $p = .037$, $\eta = .058$.