**Article:**

Kirk, Elizabeth orcid.org/0000-0002-2507-1185 and Lewis, Carine (2016) Gesture Facilitates Children's Creative Thinking. Psychological Science. pp. 225-232. ISSN 1467-9280

https://doi.org/10.1177/0956797616679183

This is a pre-publication copy of a manuscript accepted for publication in *Psychological Science*. Title: Gesture Facilitates Children's Creative Thinking. Authors*:* Elizabeth Kirk (University of York); Carine Lewis (University of Hertfordshire). Corresponding author: Elizabeth Kirk, email: elizabeth.kirk@york.ac.uk.

# Gesture Facilitates Children's Creative Thinking

Elizabeth Kirk

University of York

Carine Lewis

University of Hertfordshire

Author Note

Elizabeth Kirk, Department of Psychology, University of York; Carine Lewis, Department of Psychology, University of Hertfordshire.

Correspondence concerning this article should be addressed to Elizabeth Kirk, Department of Psychology, University of York, York, YO10 5DD. Email: elizabeth.kirk@york.ac.uk

Word count = 1933 (includes introductory and discussion material in the main text, footnotes, acknowledgments, and appendices).

**Abstract**

Gestures help people think and can help offer new ideas to problem solvers. We conducted two experiments exploring the self-oriented function of gesture in a novel domain; creative thinking. In Experiment 1 we explored the relationship between children's spontaneous gesture production and their ability to generate novel uses for everyday items (Alternative Uses Task). There was a significant correlation between children's creative fluency and their gesture production, with the majority of children's gestures depicting an action upon the target object. Restricting children from gesturing did not significantly reduce their fluency. In Experiment 2 we encouraged children to gesture and this significantly boosted their creative idea generation. These findings demonstrate that gestures serve an important self-oriented function and can assist creative thinking.

**The Role of Gesture in Children's Creative Thinking**

Gesturing helps children think. The gestures that accompany children's spoken explanations of problems can reveal understanding not found in speech and can predict knowledge change (for a review see Goldin-Meadow & Alibali, 2013). Gesture can also help create ideas. By gesturing while problem solving, new ideas and strategies are generated (Alibali & Kita, 2010; Alibali, Spencer, Knox, & Kita, 2011; Beilock & Goldin-Meadow, 2010). Evidence demonstrates that gestures highlight perceptual-motor information, thus making this information salient to the problem solver. Alibali and Kita (2010) compared children's explanations when they were free to move their hands compared to when their gestures were suppressed and found that children were more likely to focus on perceptual features of the task when they could gesture. By manipulating gesture the researchers were able to influence the child's approach to problem solving.

Depending on the nature of the problem to be solved, gesture may facilitate or hinder solution generation. For some tasks, a focus on perceptual-motor information can be detrimental (e.g. Piagetian conservation tasks and gear movement tasks, Alibali & Kita, 2010; Alibali et al., 2011). For other tasks this focus may benefit the problem solver. One such task is the Alternative Uses Task (AUT; Guilford, 1967), a measure of divergent thinking in which participants are asked to generate novel uses for everyday objects (e.g. a newspaper could be used as a fly swat). Considering the perceptual features and motor properties of the items is an effective strategy to generate novel affordances (Gilhooly, Fioratou, Anthony, & Wynn, 2007).

There is only one study to our knowledge that has investigated the importance of action for children's performance on the AUT. Dansky and Silverman (1973) tested children aged 4 to 6 years and allocated them randomly to one of three conditions; play, imitation, or control. In the play condition, children were given an opportunity to play with a set of objects (e.g. an empty matchbox) and after these were removed from view, children were asked to generate alternative uses for those objects. In the imitation condition, children observed the experimenter demonstrate a typical activity with the object (e.g. turning screws with a screwdriver) and were instructed to imitate that action before completing the AUT. Children in the control condition did not have any experience with the objects. Children who had played with the objects generated significantly more responses for each object than participants in either the imitation or control conditions. Interestingly, there was little difference in the AUT scores of children in the imitation and control conditions. Even though children in the imitation condition had physically interacted with the items, doing so did not confer the benefits that free play did. Physically manipulating the items in a playful, symbolic manner on the other hand increased the range of associations that children were able to generate. Imitating the typical action upon the item may have highlighted some perceptual-motor features, however these would have been restricted to the concrete example, thus limiting children's ability to think divergently.

This explanation is in keeping with findings from a study that compared the benefits to learning of gesture versus action. Novack, Congdon, Hermani-Lopez, and Goldin-Meadow (2014) taught children different strategies to solve mathematical equivalence problems that used different hand movements; either an action, concrete gesture, or an abstract gesture. All children learnt how to solve the problems, regardless of which strategy they had used. However, it was only those children who had been trained to use abstract gesture that were able to transfer their learning successfully to solve new problems. Just like the children in the Dansky and Silverman study, these results show that imitating an action limits children's thinking to a particular problem. On the other hand, the abstract nature of gesture frees up children to think beyond the concrete here and now.

In the context of the AUT, gestures could provide children with the possibility to symbolically represent everyday items in unlimited forms and functions, boosting the child's ability to explore alternative affordances. No studies to our knowledge have documented what children do with their hands when they complete the AUT. Thus, our first aim was to explore whether children gestured spontaneously on the task and whether doing so was correlated with the number of alternative uses they were able to generate (fluency). We predicted there would be a positive correlation between gesture and fluency.

Additionally, we tested the impact of suppressing gesture on children's ability to generate novel solutions. Alibali and Kita (2010) reported that restricting children from gesturing caused them to focus less on perceptual-motor features of the problem. Because performance on the AUT is benefited from attention to these features, we would expect children who are restricted from gesturing to generate fewer novel affordances. We tested the impact of artificially restricting children's ability to gesture by asking the same children to complete the AUT under conditions that allowed and restricted gesture. Previous research has found that the effects of experimentally suppressing gesture were identical to the effects observed when participants did not gesture spontaneously (Goldin-Meadow et al., 2001). Thus, we hypothesised that children who do not gesture (either by choice or instruction) would generate significantly fewer novel uses than children who do gesture.

## Experiment 1

### Method

***Participants.*** A sample of 78 children (47 females) was recruited to participate from schools in [location removed for blind review]. Children were aged 9-11 years ($M_{\text{age}}$ = 10.10, $SD_{age}$ = 0.56). We targeted this age range as previous research identified a peak in divergent thinking between grades 3 – 5 (Charles & Runco, 2001; Kim, 2011).

***Materials.*** The Alternative Uses Task (AUT; Guilford, 1967) was administered to children. Two sets of six everyday items were selected and matched on imageability and concreteness (norms obtained from the MRC database, Wilson, 1988; see Appendix). Manipulability norms were also obtained and are reported in the results. Items were presented individually as colour images on an A4 sized sheet of paper. Set A included: newspaper, button, tin, knife, shoe, and key. Set B included barrel, ball, pencil, bowl, kettle, and pin.

***Procedure***. A researcher visited the school and tested children individually in a quiet area of the school or classroom and administered the AUT. Children were presented with images individually and asked to generate as many novel uses as they could, with no time limit. Prompts were used to encourage children to continue responding (e.g. "What else could you do with it?").

A subsample of the children completed the AUT just once, under this control condition only ($n$ = 26) in which they were free to move their hands (gesture allowed). The remaining children ($n$ = 52[1]) completed the AUT twice; once under this control condition (gesture allowed) and again under a gesture-restricted condition (within-subjects, order of conditions counterbalanced across children). The same two word sets were used (Set A and Set B, counterbalanced). In the gesture-restricted condition, children were asked to place their hands in mittens, which were then secured with Velcro to a board on the table in front of them. Children were instructed to keep their hands still while they completed the task. Full ethical approval was granted by the (removed for blind review) ethics committee.

*AUT Coding.* All sessions were transcribed and the number of valid novel uses generated was calculated to provide a fluency score per child. For a use to be considered valid it needed to be different from the typical use. For example, reading a newspaper was not considered a valid alternative use, but using it for hitting flies was. A second coder

[1] a priori power analysis conducted using G*Power. With an $\alpha$ = .05, power = .80, sample size needed for effect size $dz$ = .5 was $N$ = 34 (between subjects t-test, 2-tailed).

independently coded 20% of the sample (*n* = 15). Inter-rater reliability was good, ICC = .91, 95% CI [.74, .97]. Originality of responses was coded by calculating the number of participants who gave any one idea by the number of participants in the sample who had received that test item. A total average originality score was then calculated by summing the scores for the individual items for each participant, and then dividing by fluency (Runco et al., 1987). Flexibility was coded by counting the number of different categories of responses to each item (example categories include arts and crafts, weapons, and a form of shelter). An average flexibility score was calculated by summing the flexibility scores across all items for each child, then dividing by fluency.

*Gesture Coding.* Children's gesture production was coded using the Observer, a computer-aided coding system. All iconic gestures were coded, which are gestures "that in form and manner of execution exhibits a meaning relevant to the simultaneously expressed linguistic meaning" (McNeil, 1985, p.354). These movements convey semantic meaning, for instance performing an action as if to roll up a newspaper with two hands. Gestures were additionally coded as one of four categories; *target item manipulation* gestures are performed in the first-person perspective and depict the performance of an action upon the target object, e.g. appearing to roll-up a newspaper with the hands; *spatial* gestures depict spatial information, including shape, size or trajectory of movement, for example depicting a large shoe by spacing hands far apart; *body part as object (BPO)* gestures represent an object with the hands, for example using a flat hand to represent a sheet of newspaper; *observer viewpoint* gestures depict information from a third person perspective, for example using the fingers to represent somebody walking upon a barrel; *other iconic* gestures depict semantic information, not including direct manipulations of the target item, for example the action of squeezing a sandwich. Each hand of a two-handed gesture (*N* = 45) was coded separately if the two hands performed different functions (for example, the left hand may represent the target object as a body part as object gesture while the right hand performs an action upon this, presenting a target item manipulation gesture). To rule out the possibility that coding two-handed gestures in this way altered the results, analyses were repeated coding two-handed gestures by their dominant function. Results were unchanged and are reported in the Supplementary Materials. To establish inter-rater reliability, 20% of the gestures were selected randomly ($n$ = 123) and coded blind by a second independent coder for gesture type. Agreement on gesture type was substantial (Cohen's $k$ = .67).

**Results**

Fluency and gesture scores were not normally distributed and transformation did not improve normality. Four outliers were identified for fluency scores and when these cases were removed fluency data were normally distributed. Three of the outliers had completed the task under both conditions (gesture allowed and gesture restricted) and one had completed the task under the gesture allowed condition only. To account for the non-normal distribution of gesture scores, bootstrap confidence intervals are reported. The mean number of valid uses that children generated (in total) was 13.64 ($SD$ = 5.75). Originality scores ($M$ = .87, $SD$ = .05) were normally distributed, however flexibility scores ($M$ = .93, $SD$ = .07) were not normally distributed. Mean gesture production is reported in Table 1, along with correlations with fluency scores (when free to gesture). Only two cases of OVP gestures were observed and so these were omitted from analyses. Total production of gesture, target item manipulation gestures and other iconic gestures all correlated significantly with fluency scores and had confidence intervals that did not include zero. Originality and Flexibility did not correlate with gesture ($p$ = .24, $p$ = .83 respectively).

Table 1

*Descriptive statistics for gesture type and association with fluency (N = 74)*

| | Frequency | Mean (SD) | Pearson correlation with fluency | Bootstrapped 95% CI |
|---|---|---|---|---|
| Total gestures | 502 | 6.78 *(12.80)* | .34 (*p* =.003)** | [0.05, 0.54] |
| Target item manipulation | 231 | 3.12 *(5.46)* | .35 (*p* =.002)** | [0.09, 0.57] |
| Other iconic | 81 | 1.09 *(3.29)* | .35 (*p* = .002)** | [.06, 0.54] |
| Spatial gestures | 116 | 1.57 *(2.66)* | .25 (*p* = .032)* | [-0.05, 0.50] |
| BPO | 72 | 0.97 *(2.53)* | .25 (*p* = .035)* | [-0.70, 0.46] |
| OVP | 2 | 0.03 *(0.23)* | - | - |

*Note*. ** significant at $p$ <.01 * significant at $p$ <.05 (2-tailed); CI = confidence interval.

*Children who did not gesture by choice.* Next, we considered children's performance in the gesture allowed condition only and compared children who spontaneously did and did not gesture. Children who produced at least one gesture were coded as 'gesturers' ($n$ = 47) and children who did not gesture were coded as 'non-gesturers' ($n$ = 27). On average gesturers generated more novel uses than non-gesturers; the mean fluency scores of gesturers was 10.60 (*SD* = 5.44) and the mean fluency of non-gesturers was 8.00 (*SD* = 5.91), however this difference did not reach significance $t(72) = -1.92$, $p = .059$. There was no impact of gesture on originality scores (gesturer *M* = .89, SD = .05, non-gesturer *M* = .83, *SD* = .19; t(28.22) = -1.45, $p = .159$) or flexibility scores (gesturer *M* = .93, *SD* = .08, non-gesturer *M* = .90, *SD* = .21; t(30.52) = -.68, $p = .503$).

*Children who did not gesture by instruction*. Of the subsample of the children ($n = 49$) who completed the AUT under two conditions (gesture allowed and gesture restricted), 30 produced at least one gesture in the gesture allowed condition. These children were labelled as 'gestured by choice' and their performance was compared when free to gesture and when restricted from gesturing. Children who gestured by choice produced more novel uses when free to gesture ($M = 9.57$, $SD = 8.57$) than when instructed not to gesture (M = 8.03, $SD = 7.34$), however this difference was not significant, $t(29) = 1.17$, $p = .251$. There was no effect of condition on originality scores (gesture allowed $M = .87$, $SD = .06$, gesture restricted $M = .81$, $SD = .23$, $t(29) = 1.41$, $p = .170$) or flexibility scores (gesture allowed $M = .94$, $SD = .09$, gesture restricted $M = .89$, $SD = .25$, $t(29) = 1.11$, $p = .277$)[2].

*The effect of item manipulability on gesture and fluency.* One possibility was that the correlation between gesture and creative fluency was driven by a third underlying variable. We considered whether the manipulability of the test items increased children's fluency while also independently increasing children's proclivity to gesture. Previous research has found that items that score high on manipulability are more likely to elicit gesture (Pine, Gurney & Fletcher, 2010). To test for this we obtained manipulability ratings for the target items from seven respondents (4 females, $M_{\text{age}}$ 32.00, $SD_{age} = 3.00$) via an online survey. Since these measures focus on perceptual manipulability of the items rather than experience with the items, it was deemed appropriate to collect responses from adults. AUT images were presented individually and respondents were asked three questions regarding manipulability. All responses were on a five-point likert scale, with higher ratings indicating greater manipulability. (1) *Ease to mime*: Please indicate the extent to which you could easily mime the action usually associated with this object so that any person looking at you doing this

[2] These analyses are repeated comparing all children who completed both conditions. The results are unchanged and are reported in the Supplementary Materials.

action could decide which object goes with this action (Magnié, Besson, Poncet, & Dolisi, 2003); (2) *Form manipulation*: Please indicate the degree to which the shape of the object implies how it should be used. For example, if you had never seen the object before, what is the likelihood you would pick it up and use it in its appropriate way (Wolk, Coslett, & Glosser, 2005); (3) *Graspability*: Please rate the manipulability of the object according to how easy it is to grasp and use the object with one hand (Salmon, McMullen, & Filliter, 2010). Mean manipulability scores for the 12 items are reported in the Appendix. Ease to mime and graspability scores were correlated, $r$ (10) = .63, $p$ = .027. No other scores were correlated ($p$ >.05). Since the three scales appeared to measure different dimensions of manipulability all three scores were retained.

Analyses were conducted to determine whether manipulability ratings of the items were associated with gesture production and fluency. There was a significant correlation between target item manipulation gestures and graspability, $r(10)$ = .64, $p$ = .026. There were no significant correlations between item manipulability scores and item fluency; Ease to mime $r$ = -.37, form manipulation $r$ = -.55, graspability $r$ = -.33 (all $p$s >.05). Therefore, item manipulability does not account for the relationship between gesture and fluency.

## Discussion

Children gestured spontaneously when they completed the AUT and the majority of their gestures depicted a manipulation of the target object. Gesturing had a positive impact on idea generation, with greater gesture production associated with higher fluency scores (but not originality or flexibility). The association between gesture and fluency was not explained by the manipulability of the items. However, highly graspable items were more likely to elicit gesture, and in turn gesture increased fluency. Children who did not gesture (by choice or instruction) had lower fluency scores than children that did gesture; however these differences did not reach significance. To further investigate the causal role of gesture, we manipulate gesture more directly in Experiment 2.

## Experiment 2

Our aim was to test whether encouraging children to gesture while they performed the AUT would improve their ability to generate novel uses. We hypothesised that children in a ‘gesture encouraged’ condition would perform significantly better on the AUT than children in a control condition. We included an additional control measure to account for any potential impact of children’s receptive verbal ability.

### Method

***Participants.*** Fifty-four children[3] (aged 8 -11, $M_{\text{age}}$ = 9.45, $SD_{age}$ = 0 .75) were recruited from a [location removed for blind review] Primary School (27 females).

***Procedure***. Children first completed the British Picture Vocabulary Scale: Third Edition (BPVS3, Dunn, Dunn, Whetton, & Burley, 1997 ) to provide a standardised measure of their receptive verbal ability. Children completed the AUT once, under either a gesture allowed (control) condition (*n* = 27) or a gesture encouraged condition (*n* = 27), between subjects. Children were allocated to condition sequentially, i.e. the first child to be tested was allocated to the gesture encouraged condition, the second child to the control condition, and so forth. The same materials were used as per Experiment 1, with the exception that the AUT set included all 12 of the items. In the gesture encouraged condition, children were instructed to “use your hands to show me how you could use objects in different ways” and were given examples. Children were prompted to gesture (e.g. “remember to use your hands”, “Use your hands and think what else it can be”). In the gesture allowed condition, children completed

[3] a priori power analysis conducted using G*Power. With $\alpha = .05$, power = .80, projected sample size for effect of $d = .8$ was $N = 52$ (independent samples t-test, 2 tailed).

the AUT with no special instructions and were free to move their hands. In both conditions, children were encouraged to provide answers with prompts, such as "What else?", and "Do you have any more ideas?". Additional analyses of the effect of prompts on fluency are reported in the Supplementary Materials. Fluency, originality, flexibility, and gesture production were coded as per Experiment 1. Since we were interested in the relationship between increased gesture and fluency, we only coded gesture production and not type. The sessions of ten children were transcribed and a second coder independently rated these for fluency. Inter-rater reliability was good, ICC = .89 (95% CI = .62, .97). To establish inter-rater reliability for gesture, 10 children's sessions were selected randomly and coded by a second independent coder for gesture from video. Inter-rater reliability was good, ICC = .90, (95% CI = .65, .97).

**Results**

Mean gesture production in the control condition was 13.00 (*SD* = 11.45). Only one child in the control condition did not gesture at all. The mean number of gestures produced by children in the gesture encouraged condition was 52.63 (*SD* = 22.81). Gesture scores were not normally distributed and transformation did not improve the distribution, therefore bootstrapped confidence intervals are reported. The overall mean fluency score was 31.09 (*SD* = 14.88), mean average originality was .85 (*SD* = .04) and mean flexibility was .91 (*SD* = .07). Mean BPVS standardised score was 96.93 (*SD* = 13.73). All variables were normally distributed with the exception of flexibility scores, which were negatively skewed. There were no significant correlations between BPVS scores and fluency, originality, flexibility, or gesture (*p*s >.40), thus language ability was not controlled for in subsequent analyses. There was no significant difference in BPVS score for children in the control and gesture encouraged condition.

The manipulation was successful in significantly increasing gesture production, $t(38.32) = 8.07$, $p < .001$, bootstrapped 95% CI [30.41, 49.70], *cohen's d* = 2.20. Next, we tested whether this enhancement to gesture resulted in improved fluency. The mean total fluency score of children in the control condition was 24.52 (*SD* = 10.00) and 37.67 (*SD* = 16.18) for children in the gesture encouraged condition, a significant difference; $t(43.35) = 3.59$, $p = .001$, bootstrapped 95% CI [6.18, 20.03], *cohen's d* = 0.98. There was no difference in originality scores [Control *M* = .85, *SD* = .04, Gesture Encouraged *M* = .86, *SD* = .04; $t(52) = .85$, $p = .401$] or flexibility scores [Control *M* = .91, *SD* = .07, Gesture Encouraged *M* = .90, *SD* = .06; $t(52) = -1.09$, $p = .280$].

### Discussion

As predicted, children who were encouraged to gesture produced significantly more valid novel uses than children in the control condition. Thus, it was possible to boost children's gesture rate, which improved their ability to generate novel responses.

## General Discussion

Previous research has shown that children's gestures serve a self-oriented function and facilitate convergent problem solving, for example mathematical and science tasks. The present study demonstrates that children's gestures also play a role in divergent thinking tasks that require more creative thought processes. We found that when generating novel affordances for everyday items, children spontaneously gestured, and the more they gestured the better their creative fluency. Artificially restricting gesture did not significantly hinder children's creative idea generation, whereas encouraging children to gesture significantly boosted their creative fluency. By directly manipulating gesture in Experiment 2 we demonstrate a causal role of gesture; gesture does not just reflect thought but also creates it.

*Why does encouraging gesture boost creative fluency?* An examination of the form of children's gestures revealed that when thinking about alternative uses for everyday items, children mostly performed imagined actions upon the target objects. By instructing children to use their hands to show us how they could use the objects, we effectively encouraged them to explore the praxic qualities of the items. We suggest that doing so both highlighted perceptual qualities of the items and enhanced their access to relevant action schemas. We know that both of these sources of information are used when identifying the use of familiar or unfamiliar tools, in both apraxic patients and healthy controls (Goldberg & Hagmann, 1998; Vingerhoets, Vandekerckhove, Honoré, Vandemaele, & Achten, 2011).

The association between gesture and fluency was not explained by the manipulability of the target items. We did however find a strong association between the graspability of an item and the likelihood that this item would elicit gestures to depict an action upon the object. This finding is inline with those of Chu and Kita (2016), who found that participants were more likely to gesture about an object that had a smooth surface then if it had a spiky surface. Together these findings support the action generation hypothesis; that co-thought gestures are generated from the representational use of actions. Graspability was unrelated to fluency, therefore graspability alone was not sufficient to activate novel ideas. It is the gesture that plays the crucial role; without the gesture, the perceptual qualities and relevant action schemas of the item were not made salient. We argue that instructing children to gesture boosted their ability to generate novel affordances by adding salience to aspects of the structure that would lend to divergent functions, and/or by tapping into their semantic knowledge of that object.

Gesture was related to the generation of novel ideas, but not to the originality or flexibility of these ideas. As we have argued, gestures highlight the salient features of items and trigger relevant action schemas. Thus, while helping generate ideas, these ideas are likely to be similar. Indeed, there was very little variability in originality and flexibility scores. Our results support the self-oriented function of gesture, demonstrating that gestures help children generate new ideas, however these ideas are not necessarily more unique or varied.

*Children who did not gesture.* While spontaneous and encouraged gesture was positively associated with fluency, not gesturing (by choice or instruction) did not significantly impact children's fluency scores. The study was sufficiently powered so sample size does not account for this null finding. In Experiment 1, children were asked to generate novel uses for 12 items, thus it is likely they employed multiple strategies, beyond those facilitated by gesture. Children identified as 'gesturers' (children who produced at least one gesture across all 12 items) were not consistently using a gesture strategy, but would have utilised a range of strategies, thus diluting the gesture effect. The positive correlation between gesture and fluency indicates that gesture was a good strategy to employ. Since children were clearly not relying solely on a gesture strategy in the gesture allowed condition, blocking their ability to gesture did not result in a significant drop in fluency scores, presumably because they had other strategies to use. The consistent trend towards gesturers performing better than non-gesturers suggested that a gesture strategy was more helpful. Indeed, by forcing children to adopt a gesture strategy as we did in Experiment 2, we observed a statistically significant increase in fluency.

Our findings add to the growing body of evidence demonstrating the facilitative role of gesture in thinking and have applications to the classroom. Asking children to move their hands while they think can help them tap into novel ideas; children should be encouraged to think with hands.

# References

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. . (2011). Spontaneous gestures influence strategy choices in problem solving. Psychological Science, 22(9), 1138-1144.

Alibali, M. W. K., S. . (2010). Gesture highlights perceptually present information for speakers. Gesture, 10(1), 3-28.

Beilock, S. L. G.-M., S. Gesture changes thought by grounding it in action. . Psychological Science, 21, 1605-1610.

Chu, M., & Kita, S. . (2015). Co-thought and Co-speech Gestures Are Generated by the Same Action Generation Process. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Dansky, J. L., Silverman, I, W. (1973). Effects of play on associative fluency in preschool-aged children. Developmental psychology, 9(1), 38-43.

de Ruiter, J. (2000). The production of gesture and speech. . In D. McNeill (Ed.), Language and Gesture (pp. 248–311). Cambridge: Cambridge University Press.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. . (1997). The British Picture Vocabulary Scale.

Gilhooly, K. J., Fioratou, E., Anthony, S. H., & Wynn, V. . (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. British Journal of Psychology, 98(4), 611-625.

Goldenberg G, H., S. (1998). Tool use and mechanical problem solving in apraxia. Neuropsychologia, 36, 581-589.

Goldin-Meadow, S. (2002). Constructing communication by hand. Cognitive Development, 17, 1385 - 1405.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. Annual Review of Psychology, 64, 257-283.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. . (2001). Explaining math: Gesturing lightens the load. Psychological Science, 12(6), 516-522.

Guilford, J. P. (1967). The nature of human intelligence. New York, NY: McGraw-Hill.

Magnié, M. N., Besson, M., Poncet, M., & Dolisi, C. (2003). The Snodgrass and Vanderwart set revisited: Norms for object manipulability and for pictorial ambiguity of objects, chimeric objects, and nonobjects. Journal of Clinical and Experimental Neuropsychology, 25(4), 521-560.

McNeill, D. (1985). So you think gestures are nonverbal? Psychological Review, 93(2), 350-371.

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From Action to Abstraction Using the Hands to Learn Math. Psychological Science, 25(4), 903-910.

Runco, M. A., Okuda, S. M., & Thurston, B. J. (1987). The psychometric properties of four systems for scoring divergent thinking tests. Journal of Psychoeducational Assessment, 5(2), 149-156.

Salmon, J. P., McMullen, P. A., & Filliter, J. H. (2010). Norms for two types of manipulability (graspability and functional usage), familiarity, and age of acquisition for 320 photographs of objects. Behavior Research Methods,42(1), 82-95.

Vingerhoets, G., Vandekerckhove, E., Honoré, P., Vandemaele, P. and Achten, E. . (2011). Neural correlates of pantomiming familiar and unfamiliar tools: Action semantics versus mechanical problem solving? Hum. Brain Mapp, 32(905-918).

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. . (2001). Gestures maintain spatial imagery. . The American Journal of Psychology, 11(4), 591-600.

Wolk, D. A., Coslett, H. B., & Glosser, G. (2005). The role of sensory-motor information in object recognition: Evidence from category-specific visual agnosia. Brain and Language, 94(2), 131-146.

# Appendix

## AUT stimuli norms

| Set | Item | Concreteness (range 100 - 700) | Imageability (range 100-700) | Ease to mime M *(SD)* | Form manipulation M *(SD)* | Graspability M *(SD)* |
|---|---|---|---|---|---|---|
| A | Newspaper | 576 | 616 | 4.50 *(0.53)* | 2.60 *(1.27)* | 3.80 *(0.92)* |
| | Button | 613 | 580 | 2.78 *(1.30)* | 1.67 *(0.87)* | 4.33 *(1.00)* |
| | Key | 612 | 618 | 4.75 *(0.71)* | 3.63 *(0.92)* | 4.75 *(0.46)* |
| | Shoe | 600 | 601 | 3.38 *(0.92)* | 2.88 *(1.13)* | 3.38 *(1.30)* |
| | Knife | 612 | 635 | 4.63 *(0.74)* | 4.75 *(0.46)* | 4.63 *(0.52)* |
| | Tin | 593 | 532 | 2.00 *(0.58)* | 3.14 *(1.22)* | 4.14 *(0.90)* |
| *Set A Mean* | | *601* | *597* | *3.67 (0.80)* | *3.11 (0.97)* | *4.17 (0.85)* |
| B | Barrel | 590 | 602 | 1.71 *(1.11)* | 3.43 *(1.51)* | 1.29 *(0.76)* |
| | Bowl | 575 | 579 | 2.57 *(0.98)* | 4.14 *(0.90)* | 3.86 *(1.07)* |
| | Ball | 615 | 622 | 4.29 *(0.76)* | 3.86 *(1.07)* | 5.00 *(0.00)* |
| | Kettle | 602 | 594 | 3.00 *(1.52)* | 3.29 *(1.11)* | 4.29 *(0.49)* |
| | Pencil | 617 | 607 | 4.57 *(0.79)* | 3.86 *(1.46)* | 4.86 *(0.38)* |
| | Pin | 600 | 576 | 3.43 *(1.27)* | 4.43 *(0.79)* | 4.86 *(0.38)* |
| *Set B Mean* | | *600* | *597* | *3.26 (1.07)* | *3.84 (1.14)* | *4.03 (0.51)* |

**Author contributions**

Kirk and Lewis developed the study concept. Both authors contributed to the study design. Data collection was performed by Collingwood, Hills and Harry under the supervision of Kirk and Lewis. Kirk and Lewis performed the data analysis and interpretation. Kirk drafted the manuscript with contributions from Lewis. Both authors approved the final version of the manuscript for submission.

**Acknowledgments**

We thank Danielle Collingwood, Victoria Hills, Chikara Harry and Lisa Stylianou for their assistance with data collection and coding. We thank the schools and the children for their participation.

**Supplementary Materials**

**Experiment One**

**Additional analysis of two-handed gestures.**

Because we had coded each hand of a two-handed gesture separately, we re-coded the data to check whether coding two-handed gestures via the more dominant function altered the results. There were 45 instances of two-handed gestures. We recoded these according to the dominant function by considering which hand conveyed the more meaningful action. Typically, we observed that one hand would perform a body part as object gesture, or a gesture to indicate holding an item, whilst the other hand performed an action upon the target item. Therefore the hand expressing the manipulation was coded. We re-analyzed our results using these new gesture data (Table 1). There were significant correlations between fluency and all gesture types. There was no significant correlation between total gesture and flexibility [$r(73) = -.14$, $p = .232$] and no significant correlations between the separate gesture types and flexibility (all *p*s >.05). There was no significant correlation between total gesture and originality [$r(73) = .14$, $p = .242$] and no significant correlations between the separate gesture types and originality (all *p*s >.05). Overall, the results indicate that coding two-handed gestures as one-handed gestures did not alter the results.

Table S2

*Descriptive statistics for gesture type and association with fluency (N = 74)*

| | Frequency | Mean (*SD*) | Pearson correlation with fluency | Bootstrapped 95% CI |
|---|---|---|---|---|
| Total gestures | 459 | 6.20 *(11.52)* | .35 (*p* =.003)** | [0.61, 0.54] |
| Target item manipulation | 221 | 2.99 *(5.04)* | .34 (*p* =.003)** | [0.08, 0.56] |
| Other iconic | 75 | 1.01 *(3.16)* | .34 (*p* = .003)** | [.08, 0.53] |
| Spatial gestures | 115 | 1.55 *(2.66)* | .25 (*p* = .036)* | [-0.05, 0.49] |
| BPO | 46 | 0.62 *(1.73)* | .29 (*p* = .012)* | [-0.02, 0.45] |
| OVP | 2 | 0.03 *(0.23)* | - | - |

*Note*. ** significant at $p$ <.01 * significant at $p$ <.05 (2-tailed); CI = confidence interval.

**Alternative analysis of children who did not gesture by instruction**

We compared the performance of all children who completed the AUT under both conditions (gesture allowed and gesture restricted), regardless of whether they gestured when they were free to do so. Children produced fewer novel uses when restricted from gesturing ($M = 6.06$, $SD = 3.20$) than when allowed to gesture ($M = 7.04$, $SD = 4.11$), however this was not a significant difference t(48) = 1.64, *p* = .108). There was no effect of condition on originality scores (gesture allowed *M* = .85, *SD* = .14, gesture restricted *M* = .79, *SD* = .21; t(48) = 1.45, *p* = .154) nor flexibility scores (gesture allowed *M* = .93, *SD* = .17; gesture restricted *M* = .90, *SD* = .25, t(48) = .60, *p* = .555).

**Experiment Two**

**Additional analyses of the effects of prompts on children's fluency.**

One possible interpretation of the finding that children were more fluent in the gesture encouraged condition was because they were being prompted to gesture, and that these encouragements might enhance fluency independent from gesturing. To account for the possibility that the significant difference between the two conditions was a result of prompting and not gesture, we coded the number of prompts that each child received in the two conditions and tested whether prompts were related to fluency. Prompts that encouraged children to continue were classified as *general* (e.g. "*Anything else?", "What else could you do with a [item]*?") and prompts that referred to gesture (in the gesture encouraged condition) were classified as *gesture prompts* (e.g. *"Remember to use your hands", "Use your hands to show me how you could use a [item]", "Use your hands and think what else it could be used for"*.) Due to technical problems, it was not possible to code the total number of prompts in six of the videos (two gesture encouraged and four control), thus these values were replaced with the mean for that condition. The mean number of general prompts per child in the control condition was 21.13 (*SD* = 6.01), with a range of 9 - 33. The mean number of general prompts per child in the gesture encouraged condition was 24.76 (*SD* = 12.56), with a range of 6 - 53 (a non-significant difference, *p* = .184). There was no significant correlation between prompts and fluency in the control condition [*r* (26) = .09, *p* = .657] or the gesture encouraged condition [*r* (26) = .28, *p* = .157]. Therefore, there was no evidence that encouragement to continue boosted children's fluency.

Next, we considered the impact of gesture prompts. The mean number of gesture prompts was 16.52 (*SD* = 7.86), with a range of 2-36. Gesture prompts were not significantly associated with gesture rate [*r*(26) = -.04, *p* = .829] or fluency [*r*(26) = .37, *p* = .057]. In the gesture encouraged condition there was a significant correlation between gesture prompts and general prompts [*r*(26) = .48, *p* = .011], such that children who required more prompting to continue, also elicited more prompts to gesture. The total number of prompts (general + gesture) that children in the gesture encouraged condition received was calculated; *M* = 41.28, *SD* = 17.74. There was no significant correlation between total prompts and fluency [*r*(26) = .36, *p* = .063].

The near significant correlation between gesture prompts and fluency might indicate that encouragement to gesture increased fluency independently from increasing gesture rate (which was not influenced by the prompts). The prompts to gesture may have urged children to persist with the task. However, if this were the case, one would expect that other prompts by the experimenter to work harder at the task would similarly increase fluency, yet this was not the case. General prompts were not significantly associated with children's fluency in the control condition or the gesture encouraged condition. Taken together, these analyses indicate that the additional encouragement to gesture (beyond the task instructions) does not explain the difference in creative fluency between the two conditions.