

Central Lancashire Online Knowledge (CLoK)

Title	Hands of Confidence: When Gestures Increase Confidence in Spatial Problem Solving
Type	Article
URL	https://clock.uclan.ac.uk/id/eprint/46003/
DOI	https://doi.org/10.1177/17470218231164270
Date	2023
Citation	Çapan, Dicle, Furman, Reyhan, Goksun, Tilbe and Eskenazi, Terry (2023) Hands of Confidence: When Gestures Increase Confidence in Spatial Problem Solving. Quarterly Journal of Experimental Psychology. ISSN 1747-0218
Creators	Çapan, Dicle, Furman, Reyhan, Goksun, Tilbe and Eskenazi, Terry

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.1177/17470218231164270>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Hands of Confidence: When Gestures Increase Confidence in Spatial Problem Solving

Journal:	<i>Quarterly Journal of Experimental Psychology</i>
Manuscript ID	QJE-STD-22-353.R2
Manuscript Type:	Standard Article
Date Submitted by the Author:	24-Feb-2023
Complete List of Authors:	Çapan, Dicle; Koc University College of Social Sciences and Humanities, Psychology Furman, Reyhan; University of Central Lancashire, Psychology Goksun, Tilbe; Koc Universitesi, Eskenazi, Terry ; Koc University College of Social Sciences and Humanities, Psychology
Keywords:	gestures, metacognition, confidence, mental rotation, spatial thinking

SCHOLARONE™
Manuscripts

Hands of Confidence: When Gestures Increase Confidence in Spatial Problem Solving

Çapan, D.¹, Furman, R.², Göksun, T.¹, and Eskenazi, T.¹

¹Koç University, Department of Psychology, Department of Psychology

²The University of Central Lancashire, Department of Psychology

Author Note

Correspondence concerning this article should be addressed to Dicle Çapan,
Psychology Department, Koç University, Rumelifeneri Yolu, Sariyer 34450, Istanbul,
Turkey. Email: dcapan17@ku.edu.tr

Abstract

This study examined whether the metacognitive system monitors the potential positive effects of gestures on spatial thinking. Participants ($N = 59$, 31F, $M_{age} = 21.67$) performed a mental rotation task, consisting of twenty-four problems varying in difficulty, and they evaluated their confidence in their answers to problems in either gesture or control conditions. The results revealed that performance and confidence were higher in the gesture condition, in which the participants were asked to use their gestures during problem-solving, compared to the control condition, extending the literature by evidencing gestures' role in metacognition. Yet, the effect was only evident for females, who already performed worse than males, and when the problems were difficult. Encouraging gestures adversely affected performance and confidence in males. Such results suggest that gestures selectively influence cognition and metacognition and highlight the importance of task- (i.e., difficulty) and individual-related variables (i.e., sex) in elucidating the links between gestures, confidence, and spatial thinking.

Keywords: confidence, representational gestures, mental rotation, metacognition, spatial thinking

Hands of Confidence: When Gestures Increase Confidence in Spatial Problem-Solving

Classically defined as “cognition about cognition” (Flavell & Ross, 1981), metacognition is the ability to monitor and control one’s cognitive process across contexts and time points (e.g., Allen & Armour-Thomas, 1991; Dai et al., 2018), facilitating adaptive behavior in goal-directed ways (e.g., van der Plas et al., 2022). It supports future perceptual decisions (Boldt et al., 2019), conscious awareness (Koriat, 1993), social interactions (e.g., Bahrami et al., 2010; Frith, 2012; Shea et al., 2014), and a broad range of cognitive processes such as decision-making (e.g., Yeung & Summerfield, 2012; Pleskac & Busemeyer, 2010), learning (e.g., Guggenmos et al., 2016), attention (e.g., Rummel & Meiser, 2013), and cognitive control (e.g., Fernandez-Duque et al., 2000). Metacognition pervades many aspects of experience, and as such, the ways in which it supports various cognitive processes are under constant scientific exploration. This study examined the relationship between metacognition and gesture use during spatial thinking in adults. We specifically asked whether the metacognitive system monitors gestures’ potential effects on the mental rotation aspect of spatial thinking.

We particularly focused on spatial thinking for two reasons. First, attempts at explaining how people metacognitively monitor spatial thinking are scarce at best (e.g., Ariel & Moffat, 2018; Desme et al., 2019; Thomas et al., 2012); therefore, identifying the mechanisms underlying spatial processing has a theoretical value. Second, many studies found that spontaneous and encouraged uses of gestures enhance spatial processes (e.g., in adults; Chu & Kita, 2008, 2011, 2016; Göksun et al., 2013; in children, Ehrlich et al., 2006; Ping et al., 2011; Wakefield et al., 2019), thus making spatial tasks suitable for exploring how gestures and metacognition interact. Gestures’ effects on spatial thinking are mainly explained in terms of their role in the gesturers’ thoughts (e.g., Goldin-Meadow & Beilock, 2010; Goldin-Meadow et al., 2009; Kita et al., 2017). However, the ways in which gestures

1
2
3 interact with the metacognitive system, responsible for monitoring and controlling the
4
5 cognitive processes under such change, have never been previously investigated. Here we
6
7 asked if encouraging adults to use gestures during spatial problem-solving would positively
8
9 influence their performance and metacognitive monitoring of such influence (i.e.,
10
11 performance confidence) for the first time in the literature.
12
13

14 **Metacognitive Monitoring of Spatial Thinking**

15
16 As a multi-layered construct, spatial thinking comprises mental operations involved in
17
18 encoding, visualizing, manipulating, retrieving, and reasoning about tools, objects, places,
19
20 and dynamic spatial displays (e.g., Ariel & Moffat, 2018; Hegarty, 2010; Newcombe, 2010;
21
22 Uttal & Cohen, 2012; Xie et al., 2018). Many everyday tasks that require individuals to
23
24 mentally orient themselves or objects in space, such as packing a suitcase or navigating
25
26 familiar and unfamiliar places, are part of spatial thinking processes (Charcharos et al., 2016).
27
28 Efficiency in spatial thinking is also closely related to better learning and reasoning outcomes
29
30 in science, technology, engineering, and mathematics (STEM) areas (e.g., Newcombe, 2016;
31
32 Wai et al., 2009; Uttal & Cohen, 2012), and it predicts attainment in work settings requiring
33
34 excellent spatial skills (e.g., Wai et al., 2009; Uttal & Cohen, 2012).
35
36
37
38
39

40 Metacognitive processes monitor spatial thinking and regulate the implementation of
41
42 compensatory acts toward increasing processing efficiency. Experimentally, the monitoring
43
44 aspect of metacognition is assessed in the form of confidence judgments. Confidence
45
46 judgments are explicit judgments about performance, presumably reflecting individuals'
47
48 thoughts about the degree of learning, the demands of cognitive processing (e.g., task
49
50 difficulty), or the accuracy of task performance (e.g., Cooke-Simpson & Voyer, 2007).
51
52 Metacognitive monitoring is a heuristic process (Dunlosky & Tauber, 2014) during which
53
54 theory-based or experience-based cues are considered (e.g., Koriat, 1997; Koriat & Levy-
55
56 Sadot, 1999; Nelson, 1996). Theory-based cues are individuals' naïve beliefs or perceptions
57
58
59
60

1
2
3 about their abilities (e.g., self-efficacy, Bandura, 1971; attitudes, Sidney et al., 2021).

4
5 Experience-based cues are about the information derived from internal signals during a
6
7 performance (e.g., Boldt et al., 2017; Fleming & Daw, 2017), such as the fluency with which
8
9 the information is processed, accessibility, familiarity, and relatedness of the target
10
11 information (e.g., Baranski & Petrusic, 1998; Koriat, 1997; Nelson & Narens, 1990). The
12
13 Dynamic Signal Detection Theory (Pleskac & Busemeyer, 2010) assumes confidence
14
15 judgments result from evidence accumulation. Individuals compute the probability of a
16
17 correct decision explicitly or implicitly during post-decisional processes integrating multiple
18
19 cues (i.e., perceptual and mnemonic strength, familiarity with the decision, and priors such as
20
21 the experience of success/failure) and end up with a confidence judgment (e.g., Boldt et al.,
22
23 2019; Fleming & Daw, 2017; Koriat & Goldsmith, 1996; Navajas et al., 2016; Rhodes &
24
25 Castel, 2008).

26
27
28
29
30
31 Despite its importance, metacognitive monitoring of spatial thinking has been the
32
33 main interest of very few studies, which primarily focus on age (e.g., Ariel & Moffat, 2018;
34
35 Thomas et al., 2012) or sex differences (e.g., Ariel et al., 2018; Cooke-Simpson & Voyer,
36
37 2007; Desme et al., 2019; Estes & Felker, 2012). The results of these studies favor males over
38
39 females in terms of performance and confidence and highlight confidence as a predictor of
40
41 performance differences between sexes. The results on age differences are inconclusive in
42
43 terms of confidence, with Thomas et al. (2012) suggesting increasing difficulties in
44
45 monitoring spatial performance with age and Ariel and Moffat (2018) indicating intact
46
47 monitoring ability spared from age-related declines in performance. All in all, the limited
48
49 number of studies on monitoring spatial ability focus on group comparisons and different
50
51 aspects of spatial thinking (e.g., self-perceptions of spatial ability and spatial visualization in
52
53 Ariel & Moffat, 2018; mental rotation in Cooke-Simpson et al., 2007 and Estes & Felker,
54
55 2012; visual-spatial working-memory in Thomas et al., 2012).

No previous study defined the cues people use to infer the quality of their spatial thinking. To date, studies have primarily focused on the cues people use to monitor episodic or semantic memory processes (e.g., Benjamin et al., 1998; Rhodes & Castel, 2008) or perceptual/sensory decisions (e.g., Kvidera & Koutstall, 2008). On the other hand, as summarized above, studies on metacognitive monitoring of spatial thinking focused on the age and sex differences and not inquire into the cues utilized during metacognitive monitoring (i.e., Ariel et al., 2018; Ariel & Moffat, 2018). Therefore, with the current empirical evidence, it is hard to explain whether monitoring spatial thinking is also a cue-driven heuristic process. It is unclear whether people attend to different cues across different spatial tasks or similar cues that differ in their informativeness during spatial performance monitoring (Ariel & Moffat, 2018; Dunlosky & Tauber, 2014). However, although the exact ways of monitoring spatial performance have not been clearly described, the findings compellingly suggest that people use fluency cues for generating and manipulating visuospatial representations and the perceived vividness of these representations as cues in determining their confidence in spatial tasks (e.g., Ariel & Moffat, 2018).

Gestures and Spatial Thinking

Gestures are hand movements that are used to represent and communicate information without directly changing the physical world and can be used to depict actions or objects (iconic gestures), reference a specific item, location, or trajectory (deictic gestures), or represent abstract ideas (metaphoric gestures) (e.g., McNeill, 1992; Kita et al., 2017). Gestures are frequently produced during spatial tasks because they are spatial in nature (McNeill, 1992). They externalize gesturers' thoughts about the problem at hand (e.g., Church & Goldin-Meadow, 1986), represent how people mentally visualize problems (e.g., Alibali et al., 1999), and give information about the spatial strategies used during problem-solving (e.g., Alibali et al., 2011). In this study, we specifically focused on iconic gestures

1
2
3 because previous studies suggest that people mostly use iconic gestures during spatial tasks,
4
5 and the frequency of iconic gestures increases when they solve and explain their solutions to
6
7 complex spatial visualization problems (e.g., Brooks et al., 2018; Chu & Kita, 2008, 2011;
8
9 Yang et al., 2020). Therefore, from now on, we will use the term gesture to refer specifically
10
11 to iconic gestures.
12
13

14
15
16 Gestures support various spatial processes. They aid mental rotation (in adults; Chu &
17
18 Kita, 2008, 2011, 2016; Göksun et al., 2013; in children, Ehrlich et al., 2006; Ping et al.,
19
20 2011; Wakefield et al., 2019), spatial reasoning (e.g., Chu & Kita, 2011), tracking dynamic
21
22 spatial displays (Macken & Ginns, 2014), and encoding spatial relations (Chong et al., 2013).
23
24 In this study, we examined gestures' effects on the mental rotation aspect of spatial thinking
25
26 and confidence, encouraging participants to use their gestures during spatial problem-solving.
27
28 Previous studies have mainly focused on spontaneous gestures during mental rotation, and
29
30 gesture use was manipulated mostly in child studies (i.e., Ehrlich et al., 2006; Wakefield et
31
32 al., 2019). These studies suggest that mental rotation skills are malleable and can be
33
34 improved through encouraging gestures during spatial problem-solving. Gestures increase
35
36 accuracy in mental rotation task performance, and their effect is even more pronounced than
37
38 the effect of a real action experience in which the participants practice rotating the objects in
39
40 question. In the only study conducted with adults, encouraging gestures, Chu and Kita (2011,
41
42 Experiment 2) found that the gesture-encouraged group performed better than the gesture-
43
44 allowed and gesture-prohibited groups in a mental rotation task. In all three groups,
45
46 participants produced more gestures during difficult problems compared to the easy
47
48 problems, suggesting that difficulty in spatial visualization triggers spontaneous and
49
50 encouraged uses of gestures. Chu and Kita (2011) also found that the frequency of gestures
51
52 declined over trials in the gesture-encouraged group, meaning that the participants became
53
54 better at solving problems through the help of gestures during task performance. This result
55
56
57
58
59
60

1
2
3 suggested that gestures do not only externalize spatial mental representations into hand
4
5 movements. Instead, they actively help internalize the computation of spatial transformations
6
7 during spatial visualization.
8
9

10
11 Gestures transform implicit knowledge into explicit by offloading immediate spatial
12
13 representations (e.g., Broaders et al., 2007; Cartmill et al., 2012; Hostetter & Alibali, 2008).
14
15 They also help build rich mental representations (Chu & Kita, 2011) and reduce the cognitive
16
17 load of the gesturer (e.g., Goldin-Meadow & Wagner, 2005). The Gesture- for -
18
19 Conceptualization hypothesis (Kita et al., 2017) forms a coherent framework for explaining
20
21 gestures' self-directed functions on cognition. It suggests that gestures differ from
22
23 propositional and verbal thinking; they are representational and do not directly change the
24
25 physical world (Novack & Goldin-Meadow, 2016). This representational nature of gestures
26
27 makes them more flexible and influential than actions. Gestures schematize information; they
28
29 facilitate the encoding of spatial information more than actions (So et al., 2014) through four
30
31 functions; activating, manipulating, packaging, and exploring spatial-motoric information for
32
33 thinking and speaking.
34
35
36
37
38
39

40
41 What is essential to consider in interpreting research results is that gestures do not
42
43 exert the same influence under different conditions and on different individuals (e.g., Kita &
44
45 Özyürek, 2003; Özer & Göksun, 2020). An important external source of variation pertains to
46
47 differences in the nature of the tasks. Gestures are helpful when perceptual-motor information
48
49 is critical for problem-solving (i.e., in mental rotation, Chu & Kita, 2011) and when spatial
50
51 processing is demanding but detrimental when external spatial tools are not needed (e.g.,
52
53 Alibali et al., 2011). There are also within-individual variations in response to gesture use,
54
55 and recent studies highlight the importance of discussing the benefits of gestures considering
56
57 the gesturer's cognitive skill set (for a review, Özer & Göksun, 2020). Research suggests that
58
59
60

1
2
3 individuals with the most heavily taxed cognitive resources (i.e., low spatial ability and visual
4 and spatial WM capacity) are the ones who gesture the most, showing that they may be using
5 gestures as a compensatory tool (e.g., Chu et al., 2014; Gillespie et al., 2014; Göksun et al.,
6 2013; Hostetter & Alibali, 2007; Smithson & Nicoladis, 2013). Recently Oviatt et al. (2021)
7 suggested that expertise is also a within-individual variation that changes the way people
8 gesture. Experts of a task know how to use their gestures in a way that would help them.
9 They dynamically downshift their rate of gesturing on easy tasks and upshift it on harder ones
10 that require cognitive effort. This line of evidence highlights the critical importance of
11 establishing external and internal variations in gesture use to fully explain the mechanisms of
12 self-directed influences of gestures on cognition.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

28 In summary, current evidence implies that gestures change the content of the mind of
29 gesturers' by activating and facilitating relevant mental simulations of physical movements
30 and spatial positions (e.g., Chu & Kita, 2008, 2011; Goldin-Meadow & Beilock, 2010;
31 Goldin-Meadow et al., 2009). The metacognitive system makes us monitor the content of our
32 minds all the time, with or without conscious awareness, to determine future actions and
33 behaviors (Shea et al., 2014). If gestures affect the content of the mind, they should also
34 influence how the metacognitive system monitors the content, and this question has not been
35 answered so far.
36
37
38
39
40
41
42
43
44

45 **The Current Study**

46
47
48 The current study examined how the metacognitive system monitors gestures'
49 potential facilitatory effects on the mental rotation aspect of spatial thinking. We tested
50 participants in gesture or control conditions in which they were encouraged to use gestures
51 during problem-solving or not given any instructions, respectively. We used Chu and Kita
52 (2011) 's mental rotation task, including twenty-four problems requiring participants to
53 manipulate 3D Shepard and Metzler's (1971) type objects in their minds. Participants chose
54
55
56
57
58
59
60

1
2
3 the rotated version of a particular object among the alternatives in each problem, which
4
5 varied in difficulty and provided confidence judgments after each solution.
6
7

8 Gestures and metacognitive processes can interact dynamically and reciprocally in
9
10 several ways during mental rotation. If gestures help maintain pre-existing spatial-motoric
11
12 mental representations and activate new ones (Kita et al., 2017), they should increase the
13
14 fluency with which information is processed. Relatedly, availability cues should be high
15
16 when confidence in the stimuli learned using gestures is evaluated. Fluency and availability
17
18 cues are the most evidenced cues on which people base their confidence. Here we speculate
19
20 that gestures influence cognition by facilitating such metacognitive cues to an extent,
21
22 attributing to metacognitive monitoring a mediatory role, especially in spatial tasks.
23
24 Relatedly, we suggest that metacognitive confidence evaluations inform the cognitive system
25
26 in turn and regulate the use of compensatory tools to increase performance (i.e., gesture use).
27
28
29

30 Based on such reasoning, we first hypothesized to observe higher mental rotation task
31
32 performance in the gesture group than in the control group because of the encouraged use of
33
34 gestures during problem-solving. Second, we expected that the effect of gestures on mental
35
36 rotation task performance would be reflected in the confidence judgments made by the
37
38 gesture group participants. We predicted that the gesture group participants would feel more
39
40 confident about their decisions than the participants in the control group, further documenting
41
42 gestures' effect on metacognition. Third, gestures were expected to be used purposefully by
43
44 the participants. We hypothesized that participants would produce more gestures when they
45
46 feel less confident (i.e., when the problems are more difficult than easy), in line with our
47
48 argument that the metacognitive system would monitor the effects of gestures during spatial
49
50 thinking and strategically regulate the uses of gestures.
51
52
53

54 Peripherally, we hypothesized that participants would show individual-related
55
56 variations in the mental rotation task performance and confidence. We mainly expected sex
57
58
59
60

1
2
3 differences in performance, favoring males, considering that a sex difference in mental
4
5 rotation is the most documented difference between males and females in the literature (e.g.,
6
7 Estes & Felker, 2012; Hegarty & Waller, 1995; Hegarty, 2018; Linn & Petersen, 1985;
8
9 Martens & Antonenko, 2012; Voyer et al., 1995). Relatedly, males and females were
10
11 expected to differ in how gesture encouragement was used. Individuals differ in interacting
12
13 with the environment (Uttal et al., 2013). Therefore, diverse ways of improving spatial
14
15 thinking, such as the gesture encouragement strategy we employed in this study, may not
16
17 operate the same for different populations that presumably differ at the baseline level (e.g.,
18
19 males and females; young and old adults, see for reviews Uttal et al., 2013; Özer & Göksun,
20
21 2021). Earlier research suggests that males and females use different strategies during mental
22
23 rotation (i.e., holistic vs. dynamic, e.g., Goldstein et al., 1990; Hegarty et al., 2018), and
24
25 males are generally more efficient at monitoring the efficiency of their strategy selection
26
27 (e.g., Ariel et al., 2018; Hegarty et al., 2018). We predicted that encouraging gesture use can
28
29 present, especially for females, a new strategy to solve mental rotation problems by changing
30
31 the way they manipulate spatial-motoric information in their minds and providing various
32
33 possibilities for what information to focus on (i.e., exploration function of gestures, Kita et
34
35 al., 2017). Males are generally better at spatial tasks and confident in their spatial abilities;
36
37 based on that, we expected to observe more pronounced effects of gestures on females'
38
39 performance and confidence than we would see in males.
40
41
42
43
44
45
46
47

48 **Method**

49 **Preregistration**

50
51 Before data collection, the current study's main hypotheses, planned procedure and
52
53 analyses were pre-registered via the Open Science Framework (OSF; <http://osf.io/>). We made
54
55 a minor change in the procedure of the preregistered study. We had initially planned to
56
57 collect prospective and retrospective confidence judgments during problem-solving, but we
58
59
60

1
2
3 only collected retrospective judgements. In terms of analysis, we had planned to carry out
4
5 separate repeated measures of ANOVAs to test our hypotheses. However, we decided to
6
7 analyze our data testing multilevel mixed-effect models and made changes on the OSF page
8
9 accordingly. We preregistered most of our analyses that we report here. Additional to the
10
11 preregistered analysis plan, we also conducted some exploratory analyses examining how the
12
13 difficulty of the problems affected confidence and accuracy in gesture and control groups in
14
15 interaction with participants' sex. We also tested new models considering sample
16
17 characteristics (i.e., excluding spontaneously gestured participants from the sample) and trial-
18
19 related factors. The data, explanations regarding the changes in the procedure and analyses, R
20
21 code, and additional appendices will be available on the study's OSF page upon acceptance
22
23 for publication (<https://osf.io/6e3fn/>).
24
25
26
27
28

29 **Participants**

30
31 We determined the sample size by conducting an a priori analysis using GPower
32
33 version 3.1 (Faul et al., 2007) since we initially planned to analyze our data with F-tests,
34
35 conducting separate ANOVAs. Results indicated that the required sample size to achieve
36
37 95% power for detecting a medium effect ($d = .05$) at a significance criterion of $\alpha = .05$ was
38
39 $N = 54$. The effect and planned sample size followed the previous studies that tested
40
41 individual differences (e.g., age, sex) in mental rotation task performance (e.g., Ariel &
42
43 Moffat, 2018; Voyer et al., 1995; Voyer, 2011) and that used a similar task (e.g., Chu & Kita,
44
45 2008, 2011; Göksun et al., 2013). Considering the potential attrition due to technical issues
46
47 during online Zoom meetings and our additional objectives of controlling for possible
48
49 moderating factors, we decided to increase the sample size by approximately 20% and
50
51 planned to collect data from sixty-five participants.
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

However, we analyzed our data, creating multilevel mixed-effect models. Power analysis for linear multilevel mixed models is complicated, with no rule to follow (e.g., Meteyard & Davies, 2020). What is generally recommended is to have as many sampling units as possible. Fewer sampling units point to more significant uncertainty in estimating fixed effects, random effect variances, and cross-level interactions, which is the main limitation on power (see Bell et al., 2010; Scherbaum & Ferreter, 2009; Snijders, 2005). According to Scherbaum and Ferreter (2009)'s study that they examined a range of simulation studies, enough sampling units for multilevel mixed effect models changes between 900 to 2500 data points, meaning 30-50 participants and 30-50 trials detecting an effect size $d = 0.3-0.4$ for psychological research. Westfall et al. (2014) also suggest that 30 participants performing a psychological task with 30 trials, corresponding to 900 data points, has a power of 0.25 for a small effect size (0.2) and 0.8 for a medium effect size (0.5).

In our study, we reached the planned sample size ($N = 65$, 35 female, age range: 18-35). However, we removed six participants' data from the analyses due to the technical issues experienced during the experiments (e.g., problems saving the data file or video recordings). Relatedly, we completed analyses with a sample of 59 participants ($M_{age} = 21.67$) with normal or corrected-to-normal vision and with no history of neurological disorders. Fifty-nine participants responded to a problem-solving task with 24 problems, making up 1416 sampling units, which is compatible with the suggestions (e.g., Scherbaum & Ferreter, 2009; Westfall et al., 2014) and the previous metacognition studies that used multilevel mixed effect models (e.g., Frank & Kuhlmann, 2016; Lajoie et al., 2020; Pescetelli et al., 2016; Whatley et al., 2021).

Specifically, 31 participants (18 females, 13 males) were tested in the control condition, and 28 participants (14 females, 14 males) were tested in the gesture condition.

1
2
3 Participants were recruited based on convenience sampling and via the subject pool of Koç
4
5 University, for which they were given course credits. Ethical approval was obtained from the
6
7 Institutional Review Board of Koç University (Ethics Code: 2021. 031.IRB3.012) on January
8
9 28, 2021.
10
11
12

13 **Measures**

14 ***Mental Rotation Task***

15
16 To assess mental rotation task performance and related confidence judgments, we
17
18 used Chu and Kita's stimuli set (2011) with their permission. We programmed the task
19
20 using a Python-based package, PsychoPy (Peirce et al., 2019).
21
22
23

24
25 In their well-validated task, Chu and Kita (2011) used Shepard and Metzler's (1971)
26
27 type three-dimensional objects that they created with a free 3D graphics creation software
28
29 called Blender (see www.blender.org). Each stimulus consists of two 3D objects at the top of
30
31 the screen and one at the bottom in the experimental set. The upper left and upper right
32
33 objects are mirror images of each other on the vertical axis. They are always in the canonical
34
35 position, meaning that their sides are parallel to the horizontal axis, the vertical axis, or the
36
37 axis pointing to depth. The lower object is a rotated version of one of the upper images by
38
39 four angles (60°, 120°, 240°, and 300°) around the bisector that goes through the object's
40
41 center between the horizontal and vertical axis (XY axis), the horizontal and in-depth axis
42
43 (XZ axis), and the vertical and in-depth axis (YZ axis).
44
45
46
47
48

49
50 Our experimental task consisted of 24 trials (Left vs. Right x 4 angles x 3 axes) as
51
52 Chu and Kita's (2011). Participants solved mental rotation problems by choosing the rotated
53
54 version of the lower object among the two upper images on the screen. In half of the trials,
55
56 the lower object was turned from the upper-left object, and in the remaining half, it was
57
58 turned from the upper-right object. Participants indicated their answers using their keyboard
59
60 (i.e., pressing "a" for the left image, pressing "s" for the right image). Different from Chu and

1
2
3 Kita (2011) 's study, in our task, participants evaluated their confidence in their answers after
4 each trial, using a scale from 0, "not confident at all" to 100, "very confident." (See Figure 1
5
6 for the schematic display of the task).
7
8
9

10
11 --- Insert Figure 1 about here ---
12
13
14

15 ***Spatial Problem-Solving Task***

16
17 We used a similar task to Chu and Kita's (2011) mental rotation task to assess and
18 control for the metacognitive efficiency of the participants in our analyses. The task was
19 adapted from Jost and Jansen (2020) and programmed using PsychoPy (Peirce et al., 2019).
20 We used the stimuli and the code to generate individual figures that were made available in
21 different libraries of cube figures by Peters and Battista (2008) and Jost and Jansen (2020).
22
23
24
25
26
27
28

29 In this task, participants solved spatial rotation problems. In each trial, two 3D
30 objects, mirror images of each other on the vertical axis, were presented at the top, and one
31 3D object was presented at the bottom of the screen. The lower object was a rotated version
32 of one of the upper images by four angles (45°, 90°, 135°, and 180°) around the bisector that
33 goes through the object's center between the vertical (y-axis) or the in-depth axis (z-axis). In
34 half of the trials, stimuli were rotated around the vertical axis; in the other half, they were
35 rotated around the in-depth axis. Every 15-trial included a different rotation in terms of angle
36 rotation (45°, 90°, 135°, and 180°). There were self-paced pauses between the trials, and
37 stimuli presentation was randomized across experimental sessions. Participants indicated
38 their answers using their keyboard (i.e., pressing "a" for the right image, pressing "s" for the
39 left image). After each answer, they evaluated their confidence in their solutions using a scale
40 ranging from "0" (not confident at all) to "10" (very confident).
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

58 The task consisted of practice and test phases. During the practice, participants
59 performed four trials and received feedback about their performance presented to them for
60

1
2
3 1000 ms (i.e., right vs. wrong). The main task consisted of 60 trials with two blocks, 30 trials
4
5 each. During the test phase, we gave no feedback to the participants. Instead, a fixation cross
6
7 ("+") was shown at the center of the screen for 400 ms. Stimulus images were sized 400px x
8
9 400px and were presented in a horizontal layout with a vertical shift on a black background in
10
11 both phases.
12
13
14

15 16 **Procedure**

17
18
19 Due to the COVID-19 pandemic, we collected this study's data online. We
20
21 administered PsychoPy-programmed tasks via Pavlovia (<https://pavlovia.org/>), an online
22
23 platform for remote data collection.
24
25

26
27 We collected data in two sessions. First, the Pavlovia link of the spatial problem-
28
29 solving task that we used to assess metacognitive efficiency was shared with the participants.
30
31 We wanted to control for the metacognitive efficiency scores of the participants in our
32
33 attempts to examine whether the gesture-cognition link is metacognitively monitored.
34
35 Therefore, it was essential to administer this task to all the participants first. Participants were
36
37 asked to complete the task in a quiet room before the main experimental session. They
38
39 completed the first task at a time of their choice, then contacted the experimenters for the
40
41 upcoming session.
42
43
44

45
46 One week after completing the first spatial problem-solving task assessing
47
48 metacognitive efficiency, we arranged Zoom meetings with the participants for the second
49
50 session via email, including information about the second session's procedure. Participants
51
52 were asked to be in a quiet room throughout the second session to increase efficiency in the
53
54 task and were informed about the video recordings. We tested participants in either gesture or
55
56 control conditions, and recorded videos of all, irrespective of the condition. Empirical
57
58 evidence suggests that people spontaneously produce co-thought gestures, especially when
59
60

1
2
3 they have difficulties solving spatial problems (e.g., Chu & Kita, 2011). Accordingly, we
4 recorded all participant sessions and coded spontaneous gesture use in the control group in
5 addition to the gestures produced by the participants in the gesture group. Once the Zoom
6 session started, the participants were asked permission for the video recordings. Gesture
7 group participants were instructed about gesture use during the task and were explicitly asked
8 to be in clear sight during the session. In the control group we did not want to miss any
9 possible uses of gestures out of sight. To avoid that, instructions given to control group
10 participants did not refer to any gesture use; however, they were asked to keep a comfortable
11 distance to the computer so they could be seen clearly throughout the video recordings.
12
13
14
15
16
17
18
19
20
21
22
23
24

25 After we started video recordings, we shared the Pavlovia link with the participants.
26 Participants solved 24 mental rotation problems, determining whether the lower object was
27 the same as the upper left or right object. In the gesture condition, the participants were given
28 the following instructions by the experimenter: "Try to solve the problems as accurately as
29 possible and use your hand gestures to help you do so in each trial." They were also shown
30 how to use their hands as if holding and rotating an object. In this way, participants learned
31 how using their hands can help with problem-solving. We did not give any instructions to the
32 participants tested in the control condition regarding the strategies they might use. Instead,
33 we only told them to solve the problems as accurately as possible. After each trial,
34 participants evaluated their confidence in their answers using a scale from 0 (not confident at
35 all) to 100 (very confident). There were no trials at the beginning, and we gave no feedback
36 to the participants throughout the task.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 The participants completed the first session in about 35 to 40 minutes, and the second
54 lasted approximately 30 minutes.
55
56
57
58
59
60

Gesture Coding

The first author coded the gestures produced by the participants via the video annotation software ELAN (Lausberg & Sloetjes, 2009). We used McNeill's gesture categorization (1992) to code representational gestures representing semantic information. Representational gestures can represent information by depicting concrete entities like actions or objects (iconic gestures). They may reference specific objects or locations (deictic gestures) or represent abstract ideas (metaphoric gestures). In line with the hypotheses of this study, we only coded and analyzed iconic gestures depicting actions or objects (e.g., an L-shaped gesture to represent the internal structure of the object or a curved handshape gesture to represent the entire structure).

Just after the data collection, 10 participants' data were randomly picked and coded by the first author and a second independent coder to test coding reliability. The intraclass correlation coefficient (ICC) was calculated to assess the reliability using a two-way mixed effects model. The results revealed a high degree of reliability between coders. The average measure ICC was .971 with a 95% confidence interval from .940 to .984 ($F(238, 238) = 44.256, p < .01$). The disagreements were solved through discussion, and criteria were set from the beginning to code the rest of the video files.

Results

Descriptive Statistics on Gesture Use

The participants produced 3156 gestures in 1176 trials during the main mental rotation task. Only the gesture group participants were asked to use their hands in problem-solving. Still, three participants in the control group spontaneously produced gestures, and the total number of gestures they produced was 10.36% of the total number of gestures produced by all participants in the study. The limited percentage of spontaneously produced gestures

1
2
3 and the limited number of participants who produced them made it hard to evaluate whether
4
5 spontaneous gesture use also contributed to accuracy and confidence in the task. Therefore,
6
7 we did not include spontaneous gestures produced by the participants in our analyses in
8
9 which we examined how the total number of produced gestures interacted with accuracy and
10
11 confidence. In Table 1, the descriptive statistics of gestures based on condition and sex are
12
13 summarized.
14
15

16
17
18 --- Insert Table 1 about here ---
19
20
21
22

23 **Descriptive Statistics on Metacognitive Efficiency Scores**

24
25 We calculated the participants' metacognitive sensitivity (meta d') and efficiency
26
27 (MRatio) scores based on a signal-detection theory (SDT) framework (Barett et al., 2013).
28
29 First, we quantified the metacognitive sensitivity (meta d') scores following Maniscalco and
30
31 Lau's (2014) meta- d' analysis, running the code by Lee (2019) (see
32
33 http://www.columbia.edu/~bsm2105/type2sdt/fit_meta_d_MLE.py) in MATLAB R2021a
34
35 (The MathWorks Inc., Natick, USA). This quantification allowed us to separate sensitivity
36
37 (i.e., the degree of successfully discriminating between correct and incorrect responses) and
38
39 response bias (i.e., the likelihood of a participant endorsing responses with high or low
40
41 confidence) in metacognitive performance.
42
43
44
45

46
47 Meta d' is a response bias-free measure that evaluates confidence judgments'
48
49 reliability. However, it is also on the same scale as the Type 1 (e.g., task performance)
50
51 sensitivity measure “ d' ” and may scale with performance (e.g., Fleming & Lau, 2014).
52
53 Therefore, it is more informative to estimate the metacognitive efficiency levels of the
54
55 participants relative to their task performance, especially when the goal is to compare
56
57 metacognitive monitoring across groups or conditions (e.g., Fleming & Lau, 2014; Ordin &
58
59
60

Polyanskaya, 2021). Considering this, we calculated the participants' metacognitive efficiency scores (i.e., MRatio) by dividing the meta d' to d' ratio and created an index of metacognitive performance that considers the individual level of task performance.

We could not calculate 11 participants' MRatio scores in MATLAB using the code, possibly due to the excessive rates of correct and incorrect answers or participants recurrently giving the same confidence judgments. We excluded 3 participants' data from the analyses after we detected that the first or second order hit rates or false alarm rates were extreme (<0.025 or >0.975). This decision was made following Arbusova et al. (2022) and considering that such extreme values prevent a stable estimation of SDT-based measures (Shekhar & Rahnev, 2021).

The average metacognitive efficiency scores (M-Ratio) for the spatial problem-solving task were 0.13 ($SE = .10$) for the gesture participants ($N = 26$, 13 females) and 0.20 ($SE = 0.15$) for the control group participants ($N = 19$, 12 females). A score of 1 indicates an ideal relationship between performance and confidence, and the degree of meta $d'/d' < 1$ refers to the degree to which a participant is metacognitively inefficient. We compared the metacognitive efficiency scores of the participants across conditions by conducting a one-way ANOVA. The results revealed no significant differences between the groups regarding the metacognitive efficiency scores, $F(1,44) = .133$, $p > .05$.

Main Analyses

To test our hypotheses, we created multilevel mixed-effects models with the lme4 (Bates et al., 2014) package in R (version 4.0.4, R Development Core Team, 2020). Multilevel mixed-effects models differ from the classical regression models in which only fixed effects are tested. We treat data points independently when we have only fixed effects in models. However, data is primarily nested because they are produced by the same

1
2
3 participants or are grouped by some other characteristics. Mixed-effects models incorporate
4
5 nested data structures of the data and are increasingly used in the literature (Luke, 2017).
6
7

8
9 We computed statistical significance using the *lmerTest* package (Kuznetsova et al.,
10
11 2013). We scaled continuous predictor variables following Bates et al. (2014). The aim of
12
13 this was to avoid convergence problems for fitted models and to facilitate interpretation. For
14
15 the comparisons between variables, including categorical predictors, we used *emmeans* and
16
17 *contrast* functions of the *emmeans* package in R (Lenth et al., 2019). The interactions
18
19 involving continuous predictors were tested and plotted with the *probe_interaction* function
20
21 of the *interactions* package. We visualized the interactions using the *interact_plot* and
22
23 *cat_plot* functions of the *ggplot2* package (Wickham, 2006). We compared different models'
24
25 significance by conducting chi-square tests (i.e., likelihood ratio tests) and evaluated variable
26
27 contributions to the tested models with the *Anova* function of the *car* package.
28
29
30
31

32
33 In the first line of analysis, we compared gesture and control groups. In the second
34
35 line of analysis, we focused on the gesture group only. In all models we tested, we included
36
37 random effects of subjects (i.e., some participants could be more successful in mental rotation
38
39 tasks) and trials (i.e., some trials could yield more errors and influence monitoring) to avoid
40
41 losing important information about variability within participants or within items, thus
42
43 increasing statistical power (Brown, 2021). Metacognitive efficiency scores (i.e., MRatio)
44
45 were added to the models as a control variable. In the third line of analysis, we tested new
46
47 models to understand and explain the data better, considering sample characteristics and trial-
48
49 related factors (see Supplementary Material).
50
51
52
53

54 **Analyses Comparing Gesture and Control Groups**

55
56 The analyses that we conducted to compare gesture and control groups comprised two
57
58 groups of multi-level mixed effect models: one with the accuracy (i.e., correct or incorrect
59
60

1
2
3 answers to the questions) and the other with the decision confidence judgments as the
4
5 outcome variables. For the accuracy analyses, the logistic version of the models was created,
6
7 considering that accuracy is binary as an outcome variable.
8
9

10
11 First, we created intercept-only null models for the accuracy and confidence analyses
12
13 separately, regressing the outcome variable (i.e., accuracy or confidence) on random effects
14
15 of trials and subjects. Next, we consecutively added our control variable MRatio to the
16
17 models and added fixed effect variables. We compared each model with the previous one that
18
19 included all the variables except the lastly added fixed effect to see whether the inclusion of
20
21 the new predictor makes the model significantly better at explaining the variance.
22
23
24

25 26 *Accuracy Comparisons between Gesture and Control Groups*

27

28
29 We included answers to 24 problems in the mental rotation task for accuracy analyses,
30
31 which yielded 1416 responses (24 items*59 participants). Of these, 1307 (92.30%) of the
32
33 responses were correct. The comparison between the null and last models revealed significant
34
35 improvement in the previous model's variance, $\chi^2(16) = 132.16, p < .001$. Since the results
36
37 showed improvements in each model with a new fixed effect, the last model was accepted
38
39 and reported below.
40
41
42

43
44 Our last mixed-effects logistic regression model included accuracy as the outcome
45
46 variable, group, difficulty, sex, and confidence as the fixed effects, MRatio as the control
47
48 variable, and subjects and trials as the random effects. The results revealed significant main
49
50 effects of problem difficulty, $\chi^2(1) = 4.64, p < .05$, and confidence, $\chi^2(1) = 21.26, p < .001$.
51
52 We found that easy problems predicted accuracy in answers by $1.43 \pm .66$. Higher confidence
53
54 judgments were also associated with accuracy by $0.38 \pm .08$. Group and sex interaction was
55
56 not significant, $p = .07$. Still, the interaction between problem difficulty and confidence
57
58 significantly predicted accuracy. The results revealed that difficult problems were evaluated
59
60

1
2
3 with lower confidence, and they predicted lower task accuracy by $-0.31 \pm .09$, $\chi^2(1) = 11.08$,
4
5 $p < .001$. Lastly, we observed a significant three-way interaction between group, sex, and
6
7 confidence variables, $\chi^2(1) = 6.23$, $p < .05$. Simple slope analysis revealed that for female
8
9 participants, confidence predicted accuracy both in the gesture and control groups; however,
10
11 confidence predicted accuracy only in the control group for male participants. Being female
12
13 and having low confidence in answers decreased task accuracy in both conditions by $-0.62 \pm$
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

The summarized analyses showed that the group itself does not predict task accuracy. Instead, accuracy was predicted by combining the difficulty of the problems, sex, and confidence in answers. As expected, easy problems were answered more correctly than difficult problems. Being a female was associated with lower task accuracy than being male. Confidence and accuracy increased parallel in the control group for males and females. But in the gesture group, confidence and accuracy increased in parallel only for the females. Males in the gesture group gave compatible confidence judgments across trials irrespective of their correct and incorrect answers (see Figure 2 and Table 2 for a summary of the final logistic mixed effect model's estimated coefficients and related standard errors).

--- Insert Figure 2 about here ---

--- Insert Table 2 about here ---

Confidence Comparisons between Gesture and Control Groups

The same line of analyses was also carried out, taking confidence as the outcome variable. As in the accuracy analyses, confidence was assessed by creating multilevel mixed-effects models consecutively. Test variables were added to the models as fixed factors one by one. Considering the significant improvement in the last model compared to the null model and the models that precede it, we decided to report only the last model, $\chi^2(15) = 177.21$, $p < .001$.

1
2
3 The last multi-level mixed effect model predicting confidence as the outcome variable
4 included group, difficulty, sex, and accuracy as the fixed effects, MRatio as the control
5 variable, and subjects and trials as random effects. The results revealed a significant main
6 effect of accuracy in predicting confidence judgments, $\chi^2(1) = 15.88, p < .001$, indicating that
7 the participants evaluated their correct answers with higher confidence by $1.08 \pm .27$. The
8 group and accuracy interaction were not significant, $p = .08$. We observed a significant
9 interaction between problem difficulty and accuracy, $\chi^2(1) = 7.17, p < .01$. This interaction
10 suggested that incorrectly answered difficult problems predicted decreases in confidence
11 judgments by $-.97 \pm .36$. Lastly, we observed a significant three-way interaction between
12 group, sex, and accuracy, $\chi^2(1) = 9.21, p < .01$. When compared to the correct answers,
13 incorrect answers were evaluated with lower confidence judgments by females and males in
14 the control group by $-1.93 \pm .74$. Males evaluated their correct answers similarly in both
15 gesture and control conditions. For females, a different pattern emerged. They evaluated their
16 correct answers with higher confidence in the gesture group compared to the control group
17 (see Figure 3 and Table 3 for a summary of the final logistic mixed effect model's estimated
18 coefficients and related standard errors).

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41 The results of the consecutive models created to predict confidence in answers were
42 compatible with the results we had from the accuracy analyses. In general, participants
43 evaluated their correct answers with higher confidence than their incorrect answers. Being in
44 the control group predicted lower confidence when accuracy and sex were considered. In the
45 gesture group, confidence judgments were higher than in the control group. Especially female
46 participants evaluated their correct answers more confidently if they were in the gesture
47 group.

48
49
50
51
52
53
54
55
56
57
58 --- Insert Figure 3 about here ---
59
60

--- Insert Table 3 about here ---

In these two lines of analysis, we first took accuracy and then confidence as the outcome variable. Theoretically, accuracy and confidence represent object-level and meta-level, respectively. They are assumed to be in continuous interaction during cognitive processing (e.g., Maniscalco & Lau, 2014; Mitchum & Kelley, 2010). Object-level (accuracy) informs meta-level (confidence) about performance efficiency, and meta-level changes the object-level performance by employing new ways of dealing with the task at hand. With these analyses, we aimed to see whether confidence and accuracy interact differently with variables of sex and group. The results showed that both confidence and accuracy interacted with sex and group variables in predicting the other, proving that they are in a reciprocal relationship, as theoretically suggested. Also, when we took accuracy as the outcome, we observed that confidence predicted accuracy for females in both gesture and control groups. In contrast, it predicted accuracy for males only in the control group. Analyses with confidence as the outcome showed that in the gesture group, gesture manipulation increased males' confidence in their incorrect answers, specifically extending the previous results.

Gesture Group-Only Analyses

In the second line of analyses, we tested the previously mentioned multi-level consecutive mixed-effects models with the gesture group only. The aim was to see whether the total number of gestures (i.e., gesture frequency) was significant in predicting accuracy, confidence, or both.

Accuracy in the Gesture Group

Accuracy was assessed by creating multiple mixed-effects logistic regression models as in the previous analyses. We included answers to 24 problems in the mental rotation task

1
2
3 for accuracy analyses in the gesture group from 28 participants. After we deleted the missing
4 data points list-wise, it yielded 660 responses instead of 672 (24 problems * 28 participants).
5
6
7 Of these, 512 (77.58%) of the responses were correct.
8
9

10
11 The last mixed-effect model included accuracy as the outcome variable, MRatio as
12 the control variable, gesture frequency, confidence, problem difficulty, sex as the fixed
13 effects, and subjects and trials as the random effects. In predicting accuracy in the gesture
14 group, gesture frequency was not significant, $p = .08$. The main effect of confidence was
15 significant in predicting accuracy in the gesture group, $\chi^2(1) = 21.52, p < .001$. Higher
16 confidence predicted higher accuracy by $1.32 \pm .28$. Problem difficulty was also significant in
17 predicting accuracy, $\chi^2(1) = 6.50, p < .05$. Difficult problems were associated with lower task
18 accuracy by -0.92 ± 0.36 . We observed significant interactions of confidence and problem
19 difficulty, $\chi^2(1) = 6.86, p < .01$, gesture frequency and sex, $\chi^2(1) = 3.96, p < .05$, and
20 confidence and sex variables, $\chi^2(1) = 5.47, p < .05$. The significant interaction of confidence
21 and problem difficulty showed that the participants evaluated hard problems with lower
22 confidence, and it predicted lower task accuracy by $-.47 \pm .22$. The significant interaction
23 between gesture frequency and sex showed that for males, the increased number of gestures
24 was associated with lower task accuracy by $-1.04 \pm .52$. In contrast, increasing gesture
25 production predicted higher task accuracy in females. The significant interaction of sex and
26 confidence suggested that females, in general, were less confident in their answers, and that
27 predicted lower accuracy in the task by $-1.03 \pm .44$.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 These analyses showed that the total number of gestures produced by the participants
52 during problem-solving does not predict task accuracy itself. Instead, problem difficulty, sex,
53 and confidence predicted accuracy in the gesture group as in the previous analyses. The
54 contribution of these new analyses was that they indicated a sex difference in gesture
55
56
57
58
59
60

1
2
3 production during mental rotation task performance. In general, females produced more
4
5 gestures during problem-solving than males, and it predicted accuracy in the task, meaning
6
7 that they benefited from using gestures. The reverse was true for the males. They produced
8
9 fewer gestures than females, and their increasing gesture production was associated with
10
11 lower accuracy. Males' task performance was not positively affected by gesture use. Instead,
12
13 it was worsened by this external instruction to use gestures. (See Figure 4 for the two-way
14
15 interactions and Table 4 for a summary of the final mixed effect model's estimated
16
17 coefficients and related standard errors).
18
19
20
21
22

23 --- Insert Figure 4 about here ---
24
25

26 --- Insert Table 4 about here ---
27
28

29 ***Confidence in the Gesture Group***

30
31 Lastly, we created the same multilevel mixed-effects models to predict confidence as
32
33 the outcome variable in the gesture group. A significant improvement was observed in the
34
35 last model, compared to the intercept-only null model, $\chi^2(15) = 108.22, p < .001$. The initial
36
37 model included subjects and trials as the random effects, MRatio as the control variable, and
38
39 gesture frequency, accuracy, difficulty, and sex as the fixed effects. The results revealed that
40
41 gesture frequency was not significant in predicting confidence as itself, $p = .09$. Accuracy
42
43 predicted confidence judgments in the gesture group, $\chi^2(1) = 25.84, p < .001$. Specifically,
44
45 the participants judged correct answers more confidently by $2.21 \pm .44$. The main effect of
46
47 sex was significant, $\chi^2(1) = 11.27, p < .001$, showing that males were confident in their
48
49 answers in general compared to the female participants by $2.15 \pm .64$. The interactions
50
51 between gesture frequency and difficulty and accuracy and difficulty were not significant in
52
53 predicting confidence judgments (p 's = .09, .07, respectively). The accuracy and sex
54
55 interaction significantly predicted confidence, $\chi^2(1) = 6.26, p < .05$. This result suggested that
56
57
58
59
60

1
2
3 females evaluated their incorrect answers with lower confidence than their correct answers by
4
5 -1.55 ± .62. These results enhanced previous results by showing that sex, even though it did
6
7 not predict accuracy by itself, predicted confidence in the gesture group. (See Figure 5 for the
8
9 two-way interactions and Table 5 for a summary of the final mixed effect model's estimated
10
11 coefficients and related standard errors).
12
13

14
15
16 --- Insert Figure 5 about here ---
17
18

19 20 Discussion

21
22 In this study, we investigated whether (1) mental rotation performance can be
23
24 enhanced by encouraging gesture use in participants during problem-solving and (2) the
25
26 metacognitive system would monitor this potential positive influence of gestures in spatial
27
28 thinking. With this aim, participants solved mental rotation problems and evaluated their
29
30 confidence after each solution in gesture or control conditions. We assessed our participants'
31
32 baseline metacognitive efficiency levels in a separate mental rotation test and controlled for it
33
34 in our analyses. Research has evidenced co-thought encouraged gestures' effects on
35
36 facilitating the manipulation of spatial-motoric input (i.e., mental rotation performance, Chu
37
38 & Kita, 2011; Erhlich et al., 2006); however, their influence on the monitoring of the
39
40 underlying mental states remains unexplored. This study was the first to investigate the
41
42 interactions between gestures and metacognitive monitoring of spatial thinking.
43
44
45
46

47
48 Our results showed that the participants had sufficient metacognitive monitoring
49
50 ability of spatial mental rotation. They were more confident in their correct answers than their
51
52 incorrect answers, and their answers to easy than difficult problems. Consistently, the
53
54 participants were more accurate in answering easy problems than difficult ones. These results
55
56 indicated that the participants successfully regulated their confidence judgments by
57
58 considering internal signals and task-related variations (i.e., difficulty), replicating previous
59
60

1
2
3 studies that reported correspondence between task performance and confidence judgments
4 during spatial thinking (e.g., Ariel et al., 2018; Ariel & Moffat, 2018; Cook-Simpson et al.,
5 2007; Thomas et al., 2012). We also observed the variations in performance and confidence
6 between the gesture and control groups. In the gesture group, mean performances and
7 confidence judgments were higher than in the control group. These results evidenced the
8 facilitatory role of producing gestures during problem-solving consistently with the previous
9 studies (e.g., in adults, Chu & Kita, 2008, 2011, 2014; So et al., 2014; in children, Ehrlich et
10 al., 2006), supported our hypotheses, and extended the literature in significant ways.

11
12 Specifically, we observed that cues aroused by encouraged uses of gestures influence
13 cognitive processing, possibly interacting with confidence during spatial thinking. This new
14 finding extends the literature by showing that motor system activation likely has a bottom-up
15 influence on metacognitive processes. From a perspective, this result is in line with several
16 previous studies (e.g., Alban & Kelley, 2013; Allen et al., 2016; Fleming et al., 2010, 2014;
17 Hildenbrand & Sanchez, 2022; Palser et al., 2018), reporting that haptic or proprioceptive
18 cues can influence how the metacognitive system monitors the content. For instance, Alban
19 and Kelley (2013) showed that judgments of learning (JOLs) for to-be-remembered words
20 increased as the perceptual experiences of weight during learning increased. In Palser and
21 colleagues' study (2018), people felt more confident when they were primed to move faster
22 in making their perceptual decisions than at a natural pace, and this was more of a case for
23 their incorrect responses. Several other studies also evidenced the motor system's
24 contribution to confidence judgments. Fleming et al. (2014) reported that disruption of the
25 motor system adversely affects metacognitive monitoring in perceptual discrimination, and
26 Allen et al. (2016) observed that manipulations of autonomic arousal modulate confidence on
27 a motion-discrimination task. The current study is qualitatively different from the mentioned
28 studies regarding methodology and the cognitive processing it examined. Previous studies
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 mainly focused on memory processes and perceptual decisions and examined metacognitive
4
5 monitoring of such processes during tasks that required whole-body movements by the
6
7 participants. Still, our results are parallel and indicate that proprioceptive and interoceptive
8
9 states influenced by the motor systems activation (e.g., by encouraging moving faster or
10
11 using gestures) serve as cues guiding confidence. Together such results show that
12
13 metacognition is not removed from perception and action. Instead, it is in close interaction
14
15 with them. This view adds to the traditional views of metacognition, which primarily focus
16
17 on the top-down influences on individuals' motor behavior (i.e., Baranski & Petrusic, 1998;
18
19 Palser et al., 2018), and highlights the importance of including the influences of bodily-
20
21 aroused motor information in the theoretical explanations of metacognitive processes.
22
23
24
25

26
27 When we conducted additional analyses to understand the determinants of gestures'
28
29 effects on confidence, we saw that we need to consider the interactions between individual
30
31 and task-related variations in defining gesture-related effects on cognition and metacognition,
32
33 in alignment with the recent discussions in the literature (for a review see Özer & Göksun,
34
35 2020). We observed that encouraging gestures did not enhance performance and confidence
36
37 by their mere production. Gestures facilitated task performance but not when the problems
38
39 were easy, and the participants already performed well with great confidence (i.e., in males).
40
41 Instead, gestures enhanced the accuracy and confidence of those who needed them.
42
43 Specifically, being in the gesture group was more of an advantage for females than males.
44
45 Females in the gesture group were more confident in their correct answers than females in the
46
47 control group. They felt more confident and became more accurate as they gestured, as we
48
49 hypothesized. The trend was the opposite for males, and it was unexpected. Confidence
50
51 judgments predicted the accuracy of males' mental rotation task performance only in the
52
53 control group. In the gesture group, the encouraged use of gestures increased males'
54
55
56
57
58
59
60

confidence in their incorrect answers, reversely affecting metacognitive monitoring of spatial mental rotation. Males felt less confident and became less accurate as they gestured.

There are possible explanations for these results. Empirical evidence defines the variations in visual-spatial working memory (WM) as one of the reasons underlying the sex differences in spatial abilities, especially in mental rotation (e.g., Kaufman, 2007; for a meta-analysis, Voyer et al., 2017; Wang & Carr, 2014). Consistently, in the gesture literature, gestures' compensatory role for those with the most heavily taxed resources is discussed (e.g., Chu et al., 2014; Eielts et al., 2020; Gillespie et al., 2014; Göksun et al., 2013; Hostetter & Alibali, 2007; Smithson & Nicoladis, 2013; for a review, Özer & Göksun, 2020). We did not explicitly assess the spatial WM abilities of the participants in this study, and they cannot fully account for the sex differences in spatial abilities. Still, based on the meta-analysis by Voyer and colleagues (2017) reporting consistent small magnitude differences between males and females in spatial WM, we speculate that maybe the females in our study were lower in visual-spatial WM than males and the gesture manipulation made them better at holding the spatial information in their WM by reducing the cognitive load (e.g., Goldin-Meadow & Wagner, 2005). One needs to monitor their thought processes to accurately evaluate performance while engaged in a task. However, such monitoring is hard if the task is novel or challenging, imposing a high cognitive load on the individual (van Merriënboer & Sweller, 2005). Voyer et al. (2017) suggest that the evidenced differences between males and females in visual-spatial WM abilities may make females more prone to monitoring difficulties in spatial domains. In our study, gestures probably acted as a compensatory tool for females by reducing the cognitive load and preventing them from experiencing monitoring difficulties. Reduced cognitive load paved the way for increased fluency and availability of cues with which people make metacognitive judgments, and in this way, gestures increased confidence, especially in females.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Kita et al. (2017) suggest that gestures activate, manipulate, and explore new spatial-motoric representations, and these functions are not mutually exclusive. Our study does not allow us to define the exact ways that gestures exerted their influence on spatial mental rotation. However, in line with Kita et al. (2017), we believe that gestures activated richer spatial-motoric representations and helped females manipulate the object in their minds by providing new ways of exploring it. One other possibility is that gestures changed females' strategies, making them better at encoding object parts and attending to the entire object (i.e., analytic strategy) rather than to the different pieces of an object to compare them with the response alternatives (i.e., holistic strategy). Males generally pursue an analytic strategy in solving mental rotation problems, whereas females use a holistic strategy. This difference in strategy selection has been closely linked to performance (e.g., Hegarty et al., 2018; Stieff et al., 2014; Xu & Franconeri, 2015). Among the different stages of solving mental rotation problems (i.e., processing and identifying the stimuli, rotating it, and deciding on an answer), visual rotation is crucial in predicting better spatial mental rotation performance (Boone & Hegarty, 2017). Insightfully, we believe that gestures have improved the processing in the rotation stage by changing strategies and making this process available to the participants through increased fluency and availability cues. As far as we know, there is only one study by Alibali and colleagues (2011) that previously documented gestures' influence in changing strategies (i.e., in predicting gear movements). Future studies can control for interindividual variations (i.e., spatial, visual WM) and assess strategy selections in males and females with either self-reports or eye-tracking data to determine precisely how gestures influence mental rotation performance and related confidence.

The results we obtained for males pointed to some mixed or even detrimental effects of encouraging gestures on performance and confidence measures. Recently Oviatt et al. (2021) reported that physical activity levels provide information about one's domain

1
2
3 expertise. They analyzed spontaneous gesture use by mathematics students when explaining
4
5 problem solutions and found that expertise is associated with reduced gesturing characterized
6
7 by briefer duration and slower velocity. Non-experts produced more gestures irrespective of
8
9 the difficulty level than experts. Experts produced fewer gestures in total, but the number of
10
11 iconic gestures they produced was higher than those produced by non-experts. Our study's
12
13 problem type was different (i.e., mathematical problems vs. mental rotation), but Oviatt et al.
14
15 (2017) 's findings support our results. Their findings indirectly suggest that experts have this
16
17 accurate insight into their abilities (i.e., better metacognitive monitoring) and strategically
18
19 employ the right gestures (i.e., more iconic gestures by experts) when needed. In our study,
20
21 males were already doing well, but we made them use their gestures during problem-solving,
22
23 giving them a strategy they did not need. This manipulation adversely affected their
24
25 performance. Despite this effect, males' confidence in their incorrect answers increased in the
26
27 gesture group, revealing that the encouraged use of gestures had an illusory positive effect on
28
29 their metacognitive monitoring. In one of the previous studies, Palser and colleagues (2018)
30
31 primed their participants to move faster than they would move during a perceptual decision
32
33 task and found that the participants felt more confident about their incorrect responses when
34
35 primed to move faster. They interpreted this result as that altering individuals' kinematics
36
37 adversely affects their ability to infer their confidence. Maybe our gesture encouragement
38
39 also altered males' already well-functioning problem-solving process and disrupted
40
41 performance monitoring. Our design does not permit us to explain what males experienced
42
43 when producing gestures. However, it is a valuable open question to ask in future studies that
44
45 would also deepen our understanding of the selective facilitatory effects of gestures on
46
47 cognition and metacognition. Future studies can compare males and females in different
48
49 spatial domains where we know females outperform males (i.e., object location memory,
50
51 Voyer et al., 2007) to see whether our findings related to gesture and metacognition would be
52
53
54
55
56
57
58
59
60

1
2
3 evident in different domains in the same direction. If our speculations are accurate, we should
4
5 obtain similar findings with reverse directions for males and females.
6
7

8 In this current study, we replicated sex differences in confidence and spatial thinking,
9
10 the most evidenced difference between sexes in the literature (e.g., Estes & Felker, 2012;
11
12 Hegarty et al., 2018), but most importantly, we showed that appropriate tools offered to
13
14 females could immediately influence their confidence and performance in mental rotation.
15
16 This result supports the socio-cognitive explanations of sex differences (e.g., differences in
17
18 childhood activities, spatial toy preferences, or gender-based expectations, Baenninger &
19
20 Newcombe, 1995, Gold et al., 2018; Nazareth et al., 2013) in mental rotation over the
21
22 neurological or hormonal ones (e.g., Aleman et al., 2004; Hugdahl et al., 2006). Further
23
24 studies can examine whether the relationships between gestures and confidence change based
25
26 on the theory-based factors that we know influence online assessments of performance (e.g.,
27
28 perceptions of stereotype threat and self-efficacy beliefs; Jost & Jansen, 2021). It would also
29
30 be fruitful to examine the same relationships with aging populations or neuropsychologically
31
32 impaired people performing poorly at spatial tasks and responding differently to gesture use.
33
34 Such comparisons can help us uncover confidence's role in regulating the relationships
35
36 between gestures and cognition.
37
38
39
40
41

42 One limitation of the current study is that we assessed confidence on a trial-by-trial
43
44 basis and focused on males' and females' subjective perceptions of confidence in their
45
46 answers in our main experimental task. Due to the reduced number of experimental trials, we
47
48 could not calculate our main task's reliability indices (e.g., MRatio, bias) (see Maniscalco &
49
50 Lau, 2014). Therefore, our study does not have any implications regarding gestures' effects
51
52 on increasing the monitoring efficiency of the participants in mental rotation tasks. Instead,
53
54 we showed that gesture manipulation made female participants subjectively more confident
55
56 in their answers, and this probably modulated the relationships between gestures and mental
57
58
59
60

1
2
3 rotation task performance. Future studies should elucidate whether gesture use enhances
4
5 metacognitive efficiency in spatial thinking in different populations, clarifying how long and
6
7 to what extent we see such influences on cognition and metacognition.
8
9

10 Second, in this study, we had three (two females, one male) control group participants
11
12 who spontaneously gestured during problem-solving. When we excluded them from our
13
14 analyses (see Supplementary Material 1.2. for details), group was still not significant in
15
16 predicting accuracy and confidence, but it interacted with confidence and accuracy in
17
18 predicting the other. After removing the spontaneously gestured participants from the sample,
19
20 we observed that both confidence and accuracy were higher in the gesture group, irrespective
21
22 of the participant's sex. These results suggested that spontaneously gestured participants
23
24 performed more like gesture group participants in terms of accuracy and confidence. Still, we
25
26 do not know why these participants gestured when the others did not. Considering Eielts et al.
27
28 (2020)'s findings pointing to significant influences of encouraged and spontaneous gestures
29
30 on problem-solving for individuals with a lower visual working-memory capacity, we
31
32 speculate that these participants can also be the participants with lower visual-working
33
34 memory capacity. Another possibility is that these participants have higher metacognitive
35
36 monitoring abilities, using their gestures strategically to increase performance when needed.
37
38 These speculations are yet to be tested in future studies.
39
40
41
42
43

44 Our primary aim in this study was to understand how the metacognitive system
45
46 interacts with gestures during spatial thinking. We formed our hypothesis based on the
47
48 assumption that efficient metacognitive monitoring is attained through increased accessible
49
50 cues, which we can achieve with encouraging gestures. Thus, different from Chu and Kita's
51
52 (2011), we only included a gesture-encouraged group but not a gesture-prohibited group in
53
54 this first study. Given our findings, it is valuable to conduct future studies comparing gesture-
55
56 encouraged, gesture-prohibited, and control groups in terms of task performance and
57
58
59
60

1
2
3 metacognitive monitoring to see whether prohibiting gesture use decreases confidence in
4 spatial performance by partially preventing access to metacognitive cues. Our results suggest
5 that encouraging gestures during problem-solving helped those who could benefit from
6 gesture use and made them more confident. If people use their gestures spontaneously when
7 they feel they need them, then prohibiting them from gesturing should be reflected in their
8 judgments as decreased confidence. Such findings would confirm our claims of an interplay
9 between gestures and metacognition in enhancing cognition.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Lastly, we conducted this study online. To make sure that we captured all spontaneous or encouraged uses of gestures, we asked participants to be clearly seen throughout the sessions in both conditions. During coding, we did not detect any case that made us suspicious of gestures out of sight. However, in some cases, control group participants were holding their hands under their tables and might have produced gestures we missed. Number of online studies after the pandemic have been increasing. Further studies can examine how this potential limitation might affect the interpretation of gesture studies' results comparing gesture production in online versus in-person experiments.

Conclusions

This study is the first to show that gestures interact with confidence in exerting their influence on cognitive processes. We showed that the metacognitive system monitors gestures' effects, particularly during mental rotation. Considering that interventions targeting confidence in spatial thinking could have outcomes for promoting interest and retention in spatial domains, we believe that our results have both theoretical and practical value. We observed the selective positive influence of gestures on females' performance and confidence in spatial thinking. This result highlights the importance of tailoring interventions to increase cognitive processing based on individuals' needs.

Supplementary Material

The Supplementary Material is available at: qjep.sagepub.com

Data Accessibility Statement

The data from the present experiment are publicly available at the Open Science Framework website: <https://osf.io/6e3fn/>

Acknowledgments

We thank Derin Eda Derelioğlu, Deniz Gülyurt, and Cemre Birkan for their help in data collection and coding. We thank Demet Özer and Emir Akbuğa for their contributions to planning the data analyses. We also thank Alper Yelimlieş, who helped us code the mental rotation task and calculate the metacognitive efficiency scores.

Declaration of Conflicting Interests

The Author(s) declare(s) that there is no conflict of interest.

References

- 1
2
3
4
5
6 Alban, M. W., & Kelley, C. M. (2013). Embodiment meets metamemory: Weight as a cue for
7
8 metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory,*
9
10 *and Cognition, 39*(5), 1628–1634. <https://doi.org/10.1037/a0032420>
11
12
13 Aleman, A., Bronk, E., Kessels, R. P. C., Koppeschaar, H. P. F., & van Honk, J. (2004). A
14
15 single administration of testosterone improves visuospatial ability in young women.
16
17 *Psychoneuroendocrinology, 29*(5), 612–617. <https://doi.org/10.1016/S0306->
18
19 [4530\(03\)00089-1](https://doi.org/10.1016/S0306-4530(03)00089-1).
20
21
22 Alibali, M. W., Bassok, M., Solomon, K. O., Syc, S. E., & Goldin-Meadow, S. (1999).
23
24 Illuminating mental representations through speech and gesture. *Psychological*
25
26 *Science, 10*(4), 327-333. <https://doi.org/10.1111/1467-9280.00163>
27
28
29 Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for
30
31 speakers. *Gesture, 10*(1), 3-28. <https://doi.org/10.1075/gest.10.1.02ali>
32
33
34 Alibali, M. W., Nathan, M. J., Wolfgram, M. S., Church, R. B., Jacobs, S. A., Johnson
35
36 Martinez, C., & Knuth, E. J. (2014). How teachers link ideas in mathematics
37
38 instruction using speech and gesture: A corpus analysis. *Cognition and*
39
40 *instruction, 32*(1), 65-100. <https://doi.org/10.1080/07370008.2013.858161>
41
42
43 Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence
44
45 strategy choices in problem solving. *Psychological science, 22*(9), 1138-1144.
46
47 <https://doi.org/10.1177/0956797611417722>
48
49
50 Allen, B. A., & Armour-Thomas, E. (1993). Construct validation of metacognition. *The*
51
52 *Journal of Psychology, 127*(2), 203-211.
53
54 <https://doi.org/10.1080/00223980.1993.9915555>
55
56
57 Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G.
58
59 (2016). Unexpected arousal modulates the influence of sensory noise on
60

- confidence. *Elife*, 5, e18103. <https://doi.org/10.7554/eLife.18103>
- Arbuzova, P., Maurer, L. K., & Filevich, E. (2022). Metacognitive domains are not aligned along a dimension of internal-external information source. *bioRxiv*.
<https://doi.org/10.1101/2022.05.03.490468>
- Ariel, R., & Moffat, S. D. (2018). Age-related similarities and differences in monitoring spatial cognition. *Aging, Neuropsychology, and Cognition*, 25(3), 351-377.
<https://doi.org/10.1080/13825585.2017.1305086>
- Ariel, R., Lembeck, N. A., Moffat, S., & Hertzog, C. (2018). Are there sex differences in confidence and metacognitive monitoring accuracy for every day, academic, and psychometrically measured spatial ability? *Intelligence*, 70, 42-51.
- Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences*, 7(4), 363-379. [https://doi.org/10.1016/1041-6080\(95\)90007-1](https://doi.org/10.1016/1041-6080(95)90007-1)
- Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, 329(5995), 1081-1085. Doi: 10.1126/science.11857
- Bandura, A. (1971). Vicarious and self-reinforcement processes. *The nature of reinforcement*, 228278.
- Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 929.
<https://doi.org/10.1037/0096-1523.24.3.929>
- Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological Methods*, 18(4), 535-552. <https://doi.org/10.1037/a0033268>

1
2
3 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models
4 using lme4. *arXiv preprint arXiv:1406.5823*

5
6
7 Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in
8 action. *Psychological science*, *21*(11), 1605-1610.
9
10
11
12 <https://doi.org/10.1177/0956797610385353>

13
14 Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When
15 retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental*
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Psychology: General, *127*(1), 55–68. <https://doi.org/10.1037/0096-3445.127.1.55>

Boldt, A., De Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on
sensitivity and bias in decision confidence. *Journal of experimental psychology:*
human perception and performance, *43*(8), 1520. <https://doi.org/10.1037/xhp0000404>

Boldt, A., Schiffer, A. M., Waszak, F., & Yeung, N. (2019). Confidence predictions affect
performance confidence and neural preparation in perceptual decision
making. *Scientific reports*, *9*(1), 1-17. <https://doi.org/10.1038/s41598-019-40681-9>

Boone, A. P., & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the
mental rotation process!. *Journal of Experimental Psychology: Learning, Memory,*
and Cognition, *43*(7), 1005. <https://doi.org/10.1037/xlm0000370>

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children
gesture brings out implicit knowledge and leads to learning. *Journal of Experimental*
Psychology: General, *136*(4), 539–550. <https://doi.org/10.1037/0096-3445.136.4.539>

Brooks, N. B., Barner, D., Frank, M., & Goldin-Meadow, S. (2018). The role of gesture in
supporting mental representations: The case of mental abacus arithmetic. *Cognitive*
science, *42*(2), 554-575. <https://doi.org/10.1111/cogs.12527>

Brown, V.A (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in*
Methods and Practices in Psychological Science. doi:[10.1177/2515245920960351](https://doi.org/10.1177/2515245920960351)

1
2
3 Brucker, B., de Koning, B., Rosenbaum, D., Ehlis, A. C., & Gerjets, P. (2022). The influence
4
5 of gestures and visuospatial ability during learning about movements with dynamic
6
7 visualizations—An fNIRS study. *Computers in Human Behavior*, *129*, 107151.

8
9
10 <https://doi.org/10.1016/j.chb.2021.107151>

11
12 Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2012). A word in the hand: action,
13
14 gesture and mental representation in humans and non-human primates. *Philosophical*
15
16 *Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 129-143.

17
18
19 <https://doi.org/10.1098/rstb.2011.0162>

20
21 Charcharos, C., Kokla, M., & Tomai, E. (2016). Investigating the influence of spatial
22
23 thinking in problem solving. In *Proceedings of the 19th AGILE Conference on*
24
25 *Geographic Information Science* (pp. 15-17).

26
27
28 Chong, A., Choi, B., Kwong, E., Chan, J., Chong, I., Ip, M., ... & So, W. C. (2013).

29
30 Producing gestures facilitates encoding of spatial relation. In *Proceedings of the*
31
32 *Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).

33
34
35 <https://escholarship.org/uc/item/3mg475gn>

36
37
38 Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into
39
40 the micro development of the motor strategy. *Journal of Experimental Psychology:*
41
42 *General*, *137*(4), 706–723. <https://doi.org/10.1037/a0013157>

43
44
45 Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem
46
47 solving. *Journal of Experimental Psychology: General*, *140*(1), 102–
48
49 116. <https://doi.org/10.1037/a0021790>

50
51
52 Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and
53
54 saliency of speech-accompanying gestures: The role of cognitive abilities and
55
56 empathy. *Journal of Experimental Psychology: General*, *143*(2), 694–709.

57
58
59 <https://doi.org/10.1037/a0033861>

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(2), 257. DOI:10.1037/xlm0000168
- Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23(1), 43-71.
[https://doi.org/10.1016/0010-0277\(86\)90053-3](https://doi.org/10.1016/0010-0277(86)90053-3)
- Cooke-Simpson, A., & Lenth et al., 2019, D. (2007). Confidence and gender differences on the mental rotation test. *Learning and Individual Differences*, 17, 181–186.
doi:10.1016/j.lindif.2007.03.009
- Dai, R., Thomas, A.K. & Taylor, H.A. (2018). When to look at maps in navigation: metacognitive control in environment learning. *Cogn. Research* 3, 36.
<https://doi.org/10.1186/s41235-018-0130-7>
- Desme, C. J., Alvarez-Vargas, D., & Pruden, S. M. (2019). Confidence Mediates The Sex Differences Observed In Mental Rotation Tests. *Development*, 49, 344-351. Retrieved from: <https://www.researchgate.net>
- Dunlosky, J., & Tauber, S. K. (2014). Understanding people's metacognitive judgments: An isomechanism framework and its implications for applied and theoretical research. In T. Perfect & D. S. Lindsay (Eds.), *Handbook of applied memory* (pp. 444-464). Thousand Oaks, CA: Sage.
- Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology*, 42(6), 1259–1268. <https://doi.org/10.1037/0012-1649.42.6.1259>
- Eielts, C., Pouw, W., Ouweland, K., Van Gog, T., Zwaan, R. A., & Paas, F. (2020). Co-thought gesturing supports more complex problem-solving in subjects with lower visual working-memory capacity. *Psychological Research*, 84(2), 502–513.

- 1
2
3 <https://doi.org/10.1007/s00426-018-1065-9>
4
5
6 Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation
7
8 performance. *Archives of Sexual Behavior*, 41, 557–570. doi:[10.1007/s10508-011-](https://doi.org/10.1007/s10508-011-9875-5)
9
10 [9875-5](https://doi.org/10.1007/s10508-011-9875-5)
11
12
13 Faul, F., Erdfelder, E., Lang, A.G. *et al.* (2007). G*Power 3: A flexible statistical power
14
15 analysis program for the social, behavioral, and biomedical sciences. *Behavior*
16
17 *Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
18
19
20
21 Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and
22
23 metacognitive regulation. *Consciousness and cognition*, 9(2), 288-307.
24
25 <https://doi.org/10.1006/ccog.2000.0447>
26
27
28 Flavell, J. H., & Ross, L. (Eds.). (1981). *Social cognitive development: Frontiers and*
29
30 *possible futures*. Cambridge University Press.
31
32
33 Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general
34
35 Bayesian framework for metacognitive computation. *Psychological review*, 124(1),
36
37 91. <https://doi.org/10.1037/rev0000045>
38
39
40 Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in human*
41
42 *neuroscience*, 8, 443. <https://doi.org/10.3389/fnhum.2014.00443>
43
44
45 Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective
46
47 accuracy to individual differences in brain structure. *Science*, 329, 1541–1543.
48
49 <http://dx.doi.org/10.1126/science.1191883>
50
51
52 Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs
53
54 jointly contribute to Vol. effects on metacognitive judgments. *Journal of*
55
56 *Experimental Psychology: Learning, Memory, and Cognition*, 43(5), 680–
57
58 693. <https://doi.org/10.1037/xlm0000332>
59
60

- 1
2
3 Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical*
4
5 *Transactions of the Royal Society B: Biological Sciences*, 367(1599), 2213-2223.
6
7 <https://doi.org/10.1098/rstb.2012.0123>
8
9
10 Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working
11
12 memory predicts co-speech gesture: Evidence from individual
13
14 differences. *Cognition*, 132(2), 174-180.
15
16 <https://doi.org/10.1016/j.cognition.2014.03.012>
17
18
19 Gold, A. U., Pendergast, P. M., Ormand, C. J., Budd, D. A., & Mueller, K. J. (2018).
20
21 Improving spatial thinking skills among undergraduate geology students through short
22
23 online training exercises. *International Journal of Science Education*, 40(18), 2205-
24
25 2225. <https://doi.org/10.1080/09500693.2018.1525621>
26
27
28 Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of
29
30 gesture. *Perspectives on psychological science*, 5(6), 664-674.
31
32 <https://doi.org/10.1177/1745691610388764>
33
34
35 Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in*
36
37 *cognitive sciences*, 9(5), 234-241. <https://doi.org/10.1016/j.tics.2005.03.006>
38
39
40 Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new
41
42 ideas about math. *Psychological science*, 20(3), 267-272.
43
44 <https://doi.org/10.1111/j.1467-9280.2009.02297.x>
45
46
47 Goldstein, D., Haldane, D. & Mitchell, C. Sex differences in visual-spatial ability: The role of
48
49 performance factors. *Memory & Cognition* 18, 546–550 (1990).
50
51 <https://doi.org/10.3758/BF03198487>
52
53
54 Göksun, T., Goldin-Meadow, S., Newcombe, N. *et al.* (2013). Individual differences in
55
56 mental rotation: what does gesture tell us? *Cogn Process* 14, 153–162.
57
58 <https://doi.org/10.1007/s10339-013-0549-1>
59
60

- 1
2
3
4 Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence
5
6 signals guide perceptual learning in the absence of external feedback. *Elife*, 5, e13388.
7
8 <https://doi.org/10.7554/eLife.13388>
9
10
11
12 Hegarty, M. (2010). Components of spatial intelligence. In *Psychology of learning and*
13
14 *motivation* (Vol. 52, pp. 265-297). Academic Press. [https://doi.org/10.1016/S0079-](https://doi.org/10.1016/S0079-7421(10)52007-3)
15
16 [7421\(10\)52007-3](https://doi.org/10.1016/S0079-7421(10)52007-3)
17
18
19
20 Hegarty, M. (2018). Ability and sex differences in spatial thinking: What does the mental
21
22 rotation test really measure?. *Psychon Bull Rev* 25, 1212–1219.
23
24 <https://doi.org/10.3758/s13423-017-1347-z>
25
26
27
28 Hegarty, M., Burte, H., & Boone, A. P. (2018). Individual differences in large-scale spatial
29
30 abilities and strategies. In *Handbook of behavioral and cognitive geography* (pp. 231-
31
32 246). Edward Elgar Publishing. <https://doi.org/10.4337/9781784717544.00022>
33
34
35 Hegarty, M., & Waller, D. A. (2005). Individual Differences in Spatial Abilities. In P. Shah
36
37 (Ed.) & A. Miyake, *The Cambridge Handbook of Visuospatial Thinking* (pp. 121–
38
39 169). Cambridge University Press. <https://doi.org/10.1017/CBO9780511610448.005>
40
41
42 Hildenbrand, L., Sanchez, C.A. (2022). Metacognitive accuracy across cognitive and physical
43
44 task domains. *Psychon Bull Rev*. <https://doi.org/10.3758/s13423-022-02066-4>
45
46
47 Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations
48
49 between verbal and spatial skills and gesture production. *Gesture*, 7(1), 73-95.
50
51 <https://doi.org/10.1075/gest.7.1.05hos>
52
53
54 Hostetter, A.B., Alibali, M.W. (2008). Visible embodiment: Gestures as simulated action.
55
56 *Psychonomic Bulletin & Review*, 15, 495–514.
57
58 <https://doi.org/10.3758/PBR.15.3.495>
59
60

1
2
3 Hugdahl, K., Thomsen, T., & Ersland, L. (2006). Sex differences in visuo-spatial processing:
4
5 an fMRI study of mental rotation. *Neuropsychologia*, *44*(9), 1575-1583.

6
7 <https://doi.org/10.1016/j.neuropsychologia.2006.01.026>

8
9
10 Jost, L., & Jansen, P. (2020). A novel approach to analyzing all trials in chronometric mental
11
12 rotation and description of a flexible extended library of stimuli. *Spatial Cognition &*
13
14 *Computation*, *20*(3), 234-256.

15
16
17 Jost, L., & Jansen, P. (2021). Are implicit affective evaluations related to mental rotation
18
19 performance? *Consciousness and Cognition*, *94*, 103178.

20
21 <https://doi.org/10.1016/j.concog.2021.103178>

22
23
24 Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability:
25
26 Can they be accounted for by differences in working memory
27
28 capacity?. *Intelligence*, *35*(3), 211-223. <https://doi.org/10.1016/j.intell.2006.07.009>

29
30
31 Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination
32
33 of speech and gesture reveal?: Evidence for an interface representation of spatial
34
35 thinking and speaking. *Journal of Memory and language*, *48*(1), 16-32.

36
37 [https://doi.org/10.1016/S0749-596X\(02\)00505-3](https://doi.org/10.1016/S0749-596X(02)00505-3)

38
39
40 Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and
41
42 speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*,
43
44 *124*(3), 245–266. <https://doi.org/10.1037/rev0000059>

45
46
47 Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of
48
49 knowing. *Psychological review*, *100*(4), 609.

50
51
52 Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach
53
54 to judgments of learning. *Journal of experimental psychology: General*, *126*(4), 349.

55
56 <https://doi.org/10.1037/0096-3445.126.4.349>

57
58
59 Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic
60

1
2
3 regulation of memory accuracy. *Psychological review*, 103(3), 490.

4
5 doi: [10.1037/0033-295x.103.3.490](https://doi.org/10.1037/0033-295x.103.3.490)

6
7
8 Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments:

9
10 Information-based and experience-based monitoring of one's own knowledge. In S.

11
12 Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 483–

13
14 502). The Guilford Press.

15
16
17 Kuznetsova, A., Christensen, R. H. B., & Brockhoff, P. B. (2013). Different tests on

18
19 lmer objects (of the lme4 package): Introducing the lmerTest package. In *The R User*

20
21 *Conference, useR* (p. 66).

22
23
24 Kvidera, S., & Koutstaal, W. (2008). Confidence and decision type under matched stimulus

25
26 conditions: Overconfidence in perceptual but not conceptual decisions. *Journal of*

27
28 *Behavioral Decision Making*, 21(3), 253-281. <https://doi.org/10.1002/bdm.587>

29
30
31 Lajoie, S. P., Li, S., & Zheng, J. (2021). The functional roles of metacognitive judgement and

32
33 emotion in predicting clinical reasoning performance with a computer simulated

34
35 environment. *Interactive Learning Environments*, 1-12.

36
37
38 <https://doi.org/10.1080/10494820.2021.1931347>

39
40
41
42 Lausberg, H., Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN

43
44 system. *Behavior Research Methods* 41, 841–849.

45
46
47 <https://doi.org/10.3758/BRM.41.3.841>

48
49 Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squares means. Retrieved

50
51 from <https://CRAN.R-project.org/package=emmeans>

52
53
54 Linn, M. C., & Petersen, A. C. (1985). Emergence and Characterization of Sex Differences in

55
56 Spatial Ability: A Meta-Analysis. *Child Development*, 56(6), 1479–1498.

57
58
59 <https://doi.org/10.2307/1130467>

- 1
2
3 Luke, S.G. Evaluating significance in linear mixed-effects models in R (2017). *Behav*
4
5 *Res* 49, 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>
6
7
8 Macken, L., & Ginns, P. (2014). Pointing and tracing gestures may enhance anatomy and
9
10 physiology learning. *Medical teacher*, 36(7), 596-601.
11
12 <https://doi.org/10.3109/0142159X.2014.899684>
13
14
15 Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data:
16
17 meta-d', response-specific meta-d', and the unequal variance SDT model. In *The*
18
19 *cognitive neuroscience of metacognition* (pp. 25-66). Springer, Berlin, Heidelberg.
20
21
22 Martens, J., & Antonenko, P. D. (2012). Narrowing gender-based performance gaps in virtual
23
24 environment navigation. *Computers in Human Behavior*, 28(3), 809-819.
25
26 <https://doi.org/10.1016/j.chb.2012.01.008>
27
28
29 McNeill, D. (1992). Hand and mind. Chicago and London. *The University of Chicago Press*.
30
31
32 Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time
33
34 allocation. *Journal of memory and language*, 52(4), 463-477.
35
36 <https://doi.org/10.1016/j.jml.2004.12.001>
37
38
39 Metcalfe, J., and Shimamura, A. (1994). *Metacognition: Knowing About Knowing*.
40
41 Cambridge, MA: MIT Press.
42
43
44 Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models
45
46 in psychological science. *Journal of Memory and Language*, 112, 104092.
47
48 <https://doi.org/10.1016/j.jml.2020.104092>
49
50
51
52 Mitchum, A. L., & Kelley, C. M. (2010). Solve the problem first: Constructive solution
53
54 strategies can influence the accuracy of retrospective confidence judgments. *Journal*
55
56 *of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 699.
57
58
59 Navajas, J., Bahrami, B., & Latham, P. E. (2016). Post-decisional accounts of biases in
60

confidence. *Current opinion in behavioral sciences*, 11, 55-60.

<https://doi.org/10.1016/j.cobeha.2016.05.005>

Nazareth, A., Herrera, A., & Pruden, S. M. (2013). Explaining sex differences in mental rotation: role of spatial activity experience. *Cognitive processing*, 14(2), 201-204.

<https://doi.org/10.1007/s10339-013-0542-8>

Nelson, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item comments on Schraw (1995). *Applied Cognitive Psychology*, 10(3), 257-260.

[https://doi.org/10.1002/\(SICI\)1099-0720\(199606\)10:3<257::AID-ACP400>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1099-0720(199606)10:3<257::AID-ACP400>3.0.CO;2-9)

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. *The Psychology of Learning and Motivation*. Vol. 26.

Newcombe, N. S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *American educator*, 34(2), 29

Newcombe, N. S. (2016). Thinking spatially in the science classroom. *Current Opinion in Behavioral Sciences*, 10, 1-6. <https://doi.org/10.1016/j.cobeha.2016.04.010>

Novack, M.A., Goldin-Meadow, S. Gesture as representational action: A paper about function. *Psychon Bull Rev*, 24, 652–665 (2017). <https://doi.org/10.3758/s13423-016-1145-z>

Ordin, M., & Polyanskaya, L. (2021). The role of metacognition in recognition of the content of statistical learning. *Psychonomic Bulletin & Review*, 28(1), 333-340.

<https://doi.org/10.3758/s13423-020-01800-0>

Oviatt, S., Lin, J., & Sriramulu, A. (2021). I know what you know: what hand movements reveal about domain expertise. *ACM Transactions on Interactive Intelligent Systems*, 11(1), 1-26. <https://doi.org/10.1145/3423049>

- 1
2
3 Özer, D., & Göksun, T. (2020). Gesture use and processing: A review on individual
4
5 differences in cognitive resources. *Frontiers in Psychology, 11*, 573555.
6
7 <https://doi.org/10.3389/fpsyg.2020.573555>
8
9
- 10 Palser, E. R., Fotopoulou, A., & Kilner, J. M. (2018). Altering movement parameters disrupts
11
12 metacognitive accuracy. *Consciousness and cognition, 57*, 33-40.
13
14 <https://doi.org/10.1016/j.concog.2017.11.005>
15
16
- 17 Peirce, J., Gray, J.R., Simpson, S. et al. (2019). PsychoPy2: Experiments in behavior made
18
19 easy. *Behav Res, 51*, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
20
21
- 22 Pescetelli, N., Rees, G., & Bahrami, B. (2016). The perceptual and social components of
23
24 metacognition. *Journal of Experimental Psychology: General, 145*(8), 949–965.
25
26 <https://doi.org/10.1037/xge0000180>
27
28
- 29 Ping, R., Ratliff, K., Hickey, E., & Levine, S. (2011). Using manual rotation and gesture to
30
31 improve mental rotation in preschoolers. In *Proceedings of the annual meeting of the*
32
33 *cognitive science society* (Vol. 33, No. 33).
34
35
- 36 Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of
37
38 choice, decision time, and confidence. *Psychological review, 117*(3), 864-901.
39
40 <https://doi.org/10.1037/a0019737>
41
42
- 43 Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual
44
45 information: evidence for metacognitive illusions. *Journal of experimental*
46
47 *psychology: General, 137*(4), 615. <https://doi.org/10.1037/a0013684>
48
49
- 50 Rummel, J., & Meiser, T. (2013). The role of metacognition in prospective memory:
51
52 Anticipated task demands influence attention allocation strategies. *Consciousness and*
53
54 *Cognition, 22*(3), 931-943. <https://doi.org/10.1016/j.concog.2013.06.006>
55
56
- 57 Scherbaum, C. A., & Ferrer, J. M. (2009). Estimating statistical power and required sample
58
59 sizes for organizational research using multilevel modeling. *Organizational Research*
60

- 1
2
3
4 *Methods*, 12(2), 347–367. <https://doi.org/10.1177/1094428107308906>
- 5
6
7 Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal
8
9 cognitive control and metacognition. *Trends in cognitive sciences*, 18(4), 186-193.
10
11 <https://doi.org/10.1016/j.tics.2014.01.006>
- 12
13
14 Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional
15
16 objects. *Science*, 171(3972), 701-703.
- 17
18
19 Shekhar, M., & Rahnev, D. (2021). The nature of metacognitive inefficiency in perceptual
20
21 decision making. *Psychological Review*, 128(1), 45–
22
23 70. <https://doi.org/10.1037/rev0000249>
- 24
25
26 Sidney, P. G., Thompson, C. A., Fitzsimmons, C., & Taber, J. M. (2021). Children’s and
27
28 adults’ math attitudes are differentiated by number type. *The Journal of Experimental*
29
30 *Education*, 89(1), 1-32. <https://doi.org/10.1080/00220973.2019.1653815>
- 31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Smithson, L., & Nicoladis, E. (2013). Verbal memory resources predict iconic gesture use
among monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 16(4),
934-944. <https://doi:10.1017/S1366728913000175>
- Snijders, Tom A.B. ‘Power and Sample Size in Multilevel Linear Models’. In: B.S. Everitt
and D.C. Howell (eds.), *Encyclopedia of Statistics in Behavioral Science*. Volume 3,
1570–1573. Chices- ter (etc.): Wiley, 2005.
- So, W. C., Shum, P. L. C., & Wong, M. K. Y. (2015). Gesture is more effective than spatial
language in encoding spatial information. *Quarterly Journal of Experimental*
Psychology, 68(12), 2384-2401. <https://doi.org/10.1080/17470218.2015.1015431>
- Stieff, M., Dixon, B. L., Ryu, M., Kumi, B. C., & Hegarty, M. (2014). Strategy training
eliminates sex differences in spatial problem solving in a stem domain. *Journal of*
Educational Psychology, 106(2), 390. <https://doi.org/10.1037/a0034823>
- Thomas, A. K., Bonura, B. M., Taylor, H. A., & Brunyé, T. T. (2012). Metacognitive

- 1
2
3
4 monitoring in visuospatial working memory. *Psychology and Aging*, 27(4), 1099–
5
6 1110. doi:[10.1037/a0028556](https://doi.org/10.1037/a0028556)
7
8
9 Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and
10
11 how?. In *Psychology of learning and motivation* (Vol. 57, pp. 147-181). Academic
12
13 Press. <https://doi.org/10.1016/B978-0-12-394293-7.00004-2>
14
15
16 Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., &
17
18 Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training
19
20 studies. *Psychological bulletin*, 139(2), 352-402. <https://doi.org/10.1037/a0028446>
21
22
23 van der Plas, E., Zhang, S., Dong, K., Bang, D., Li, J., Wright, N. D., & Fleming, S. M.
24
25 (2022). Identifying cultural differences in metacognition. *Journal of Experimental*
26
27 *Psychology: General*. <https://doi.org/10.1037/xge0001209>
28
29
30 van Merriënboer, J.J.G., Sweller, J. (2005). Cognitive Load Theory and Complex Learning:
31
32 Recent Developments and Future Directions. *Educ Psychol Rev* 17, 147–177.
33
34 <https://doi.org/10.1007/s10648-005-3951-0>
35
36
37 Voyer, D., Postma, A., Brake, B., & Imperato-McGinley, J. (2007). Gender differences in
38
39 object location memory: A meta-analysis. *Psychonomic Bulletin and Review*, 14, 23–
40
41 38.
42
43
44 Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial
45
46 abilities: A meta-analysis and consideration of critical variables. *Psychological*
47
48 *Bulletin*, 117, 250–270. doi:[10.1037/0033-2909.117.2.250](https://doi.org/10.1037/0033-2909.117.2.250)
49
50
51
52 Voyer, D., Voyer, S.D., & Saint-Aubin, J. (2017). Sex differences in visual-spatial working
53
54 memory: A meta-analysis. *Psychon Bull Rev* 24, 307–334.
55
56 <https://doi.org/10.3758/s13423-016-1085-7>
57
58
59 Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning
60

1
2
3 over 50 years of cumulative psychological knowledge solidifies its
4 importance. *Journal of educational Psychology*, 101(4), 817-835.

7 <https://doi.org/10.1037/a0016127>

9
10 Wakefield, E. M., Foley, A. E., Ping, R., Villarreal, J. N., Goldin-Meadow, S., & Levine, S.

11
12 C. (2019). Breaking down gesture and action in mental rotation: Understanding the
13 components of movement that promote learning. *Developmental Psychology*, 55(5),

14
15 981–993. <https://doi.org/10.1037/dev0000697>

16
17 Wang, L., & Carr, M. (2014). Working memory and strategy use contribute to gender
18 differences in spatial ability. *Educational Psychologist*, 49(4), 261-282.

19
20 <https://doi.org/10.1080/00461520.2014.960568>

21
22 Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in

23
24 experiments in which samples of participants respond to samples of stimuli. *Journal*
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
of Experimental Psychology: General, 143(5), 2020

2045. <https://doi.org/10.1037/xge0000014>

Whatley, M.C., Castel, A.D. The role of metacognition and schematic support in younger and

older adults' episodic memory. *Mem Cogn*, 50, 601–616 (2022).

<https://doi.org/10.3758/s13421-021-01169-y>

Wickham, H. (2006). An introduction to ggplot: An implementation of the grammar of

graphics in R. *Statistics*, 1-8.

Xie, J., Cheung, H., Shen, M., & Wang, R. (2018). Mental rotation in false belief

understanding. *Cognitive Science*, 42(4), 1179-1206.

<https://doi.org/10.1111/cogs.12594>

Xu, Y., & Franconeri, S. L. (2015). Capacity for visual features in mental

rotation. *Psychological science*, 26(8), 1241-1251.

<https://doi.org/10.1177/0956797615585002>

1
2
3 Yang, W., Liu, H., Chen, N., Xu, P., & Lin, X. (2020). Is early spatial skills training
4
5 effective? A meta-analysis. *Frontiers in psychology, 11*, 1938.

6
7 <https://doi.org/10.3389/fpsyg.2020.01938>

8
9
10 Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence
11
12 and error monitoring. *Philosophical Transactions of the Royal Society B: Biological*
13
14 *Sciences, 367*(1594), 1310-1321. <https://doi.org/10.1098/rstb.2011.0416>

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Peer Review Version

Figure Captions

Figure 1. Schematic Display of the Mental Rotation Task with Example Stimuli

Note. The lower object represents the rotated version of the upper-left object about 240 from the bisector of the horizontal and vertical axes. The scale represents the participants' actual scale to make their confidence judgments.

Figure 2. The Relationships Between Log Odds of Accuracy (i.e., correct responding in the task) and Scaled Scores of Confidence Judgments as Factors of Sex (i.e., females, males) and Group (i.e., gesture or control).

Note. The hues around regression lines represent 95% confidence intervals.

Figure 3. The Interaction of Group (i.e., control and gesture), Answer Type (i.e., incorrect, correct), and Sex (i.e., females, males) in Predicting Confidence in Answers

Note. The brackets show 95% confidence intervals.

Figure 4. The Relationships Between Log Odds of Accuracy (i.e., Correct Responding in Task) and Scaled Scores of Gesture Frequency as a Function of Sex

Note. The hues around regression lines represent 95% confidence intervals.

Figure 5. The Interaction of Sex (i.e., Females, Males) and Accuracy (i.e., Correct, Incorrect) in Predicting Confidence in Answers

Note. The brackets show 95% confidence intervals.

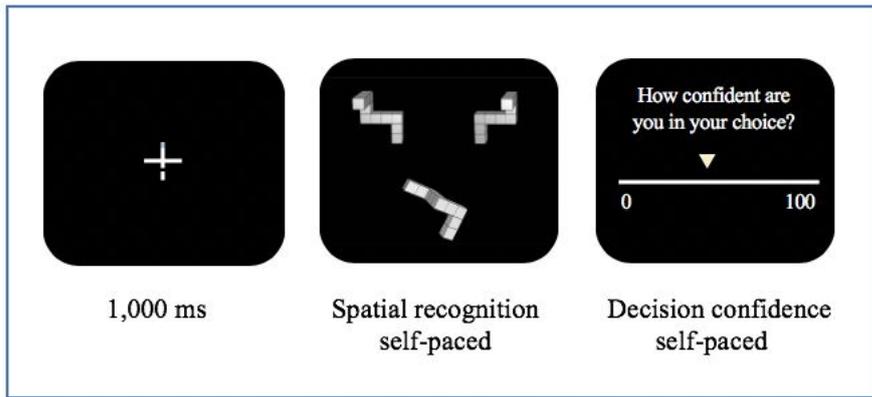


Figure 1. *Schematic Display of the Mental Rotation Task with Example Stimuli*

Note. The lower object represents the rotated version of the upper-left object about 240 from the bisector of the horizontal and vertical axes. The scale represents the participants' actual scale to make their confidence judgments.

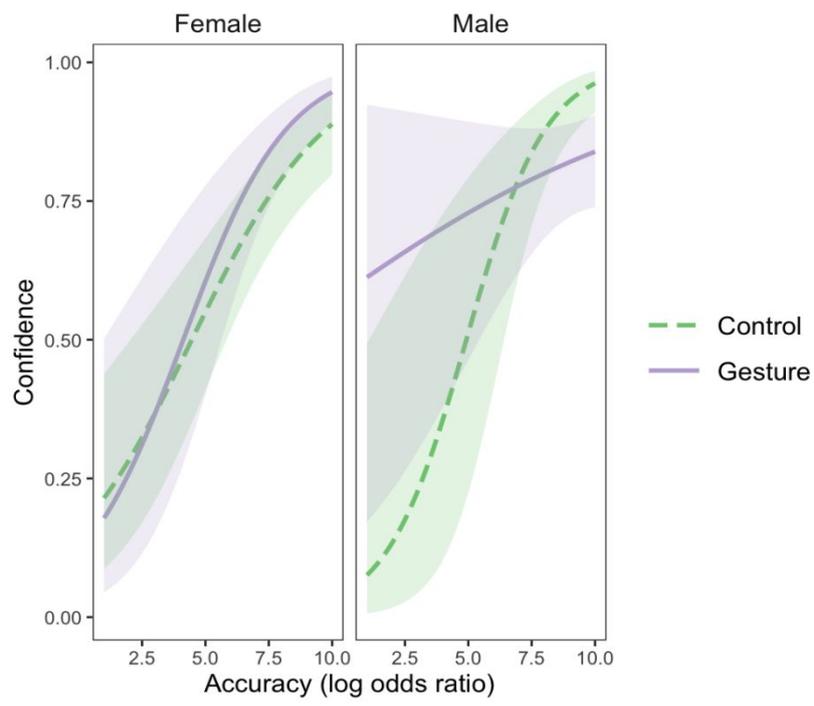


Figure 2. The Relationships Between Log Odds of Accuracy (i.e., correct responding in the task) and Scaled Scores of Confidence Judgments as Factors of Sex (i.e., females, males) and Group (i.e., gesture or control).

Note. The hues around regression lines represent 95% confidence intervals.

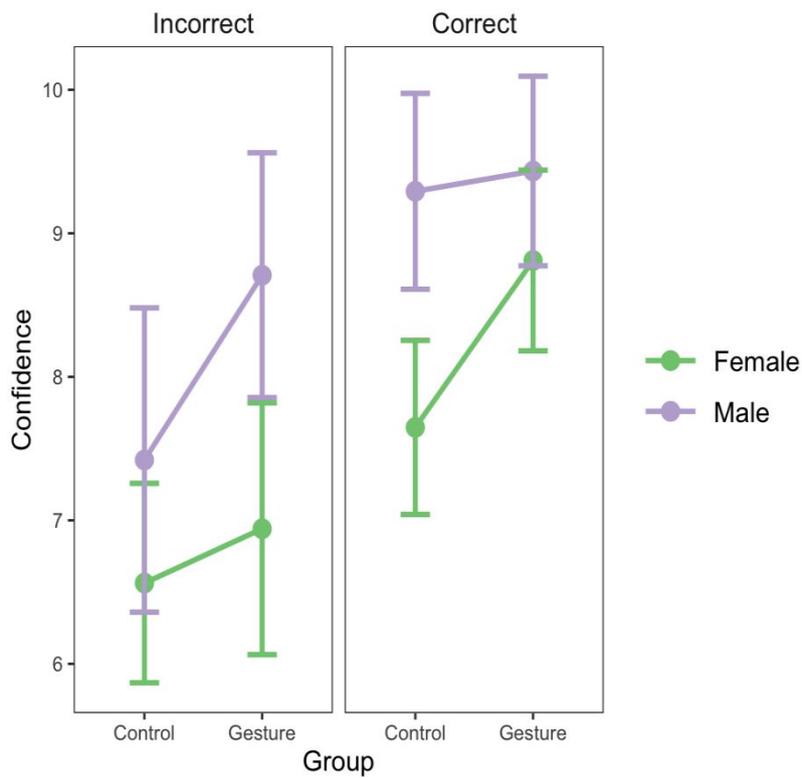


Figure 3. The Interaction of Group (i.e., control and gesture), Answer Type (i.e., incorrect, correct), and Sex (i.e., females, males) in Predicting Confidence in Answers

Note. The brackets show 95% confidence intervals.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

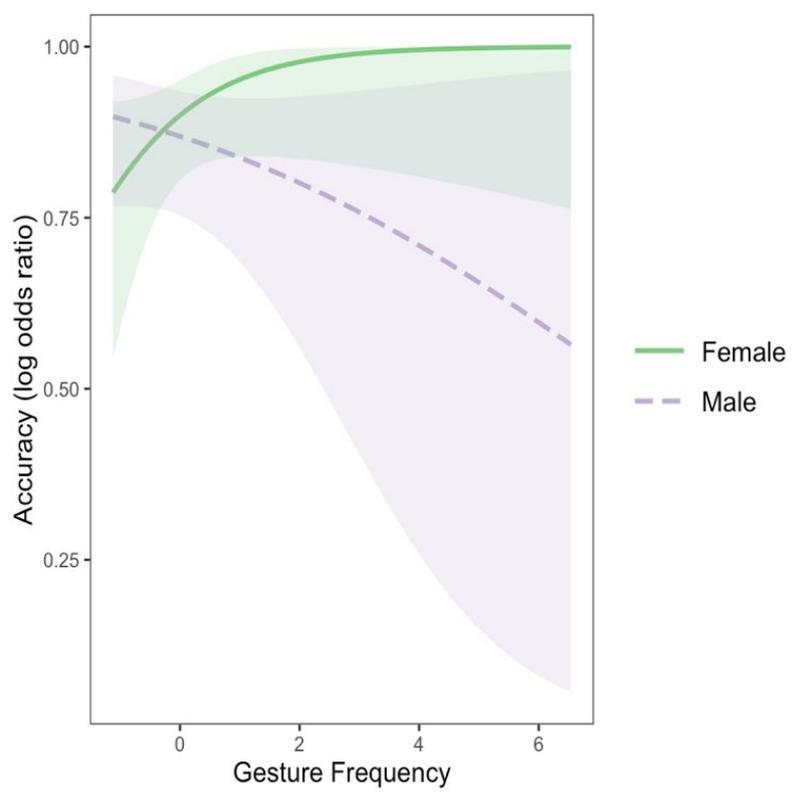


Figure 4. The Relationships Between Log Odds of Accuracy (i.e., Correct Responding in Task) and Scaled Scores of Gesture Frequency as a Function of Sex

Note. The hues around regression lines represent 95% confidence intervals.

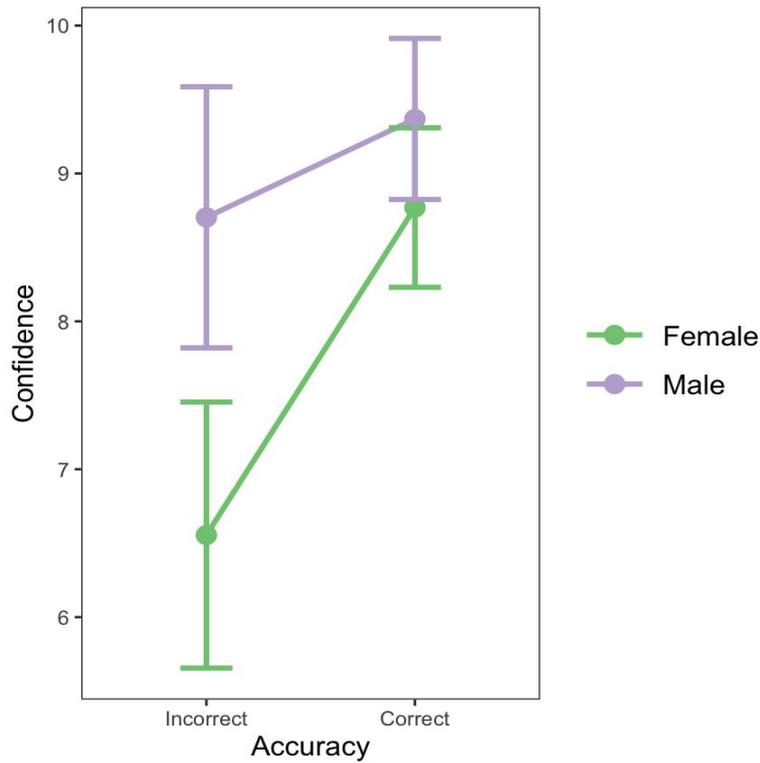


Figure 5. The Interaction of Sex (i.e., Females, Males) and Accuracy (i.e., Correct, Incorrect) in Predicting Confidence in Answers

Note. The brackets show 95% confidence intervals.

Table 1. Descriptive Statistics of Gestures as a Function of Group and Sex

Iconic Gestures	Experimental Group		
	Control	Gesture	Total
Females	201 (6.37%)	1508 (47.78%)	1709 (54.15%)
Males	126 (3.99%)	1321 (41.86%)	1447 (45.85%)
Total	327 (10.36%)	2829 (89.64%)	3156 (100%)

Note. Columns represent the total number of gestures produced by the participants in gesture and control conditions categorized by sex. The percentage of produced gestures in relation to the total number of gestures produced in the study was given in parentheses.

Table 2

Model Summary for Task Accuracy Comparing Gesture and Control Groups

	Coefficient	SE
FIXED-EFFECTS		
(Intercept)	-1.67**	.61
MRatio	-0.01	.11
Group	-0.33	1.07
Difficulty	1.43*	.66
Sex	-1.44	1.54
Confidence	0.38***	.08
Group*Difficulty	0.44	1.19
Group*Sex	3.78 †	2.12
Difficulty*Sex	0.06	1.77
Group*Confidence	0.11	.13
Difficulty*Confidence	-0.31***	.09
Sex*Confidence	0.26	.18
Group*Difficulty*Sex	-2.49	2.38
Group*Difficulty*Confidence	0.004	.16
Group*Sex*Confidence	-0.62*	.25
Difficulty*Sex*Confidence	-0.01	.21
Group*Difficulty*Sex*Confidence	0.33	.28
Variance		
RANDOM-EFFECTS		
Intercepts		
Subjects	5.21	
Trials	1.35	
<i>Note.</i> Significance codes = *** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .1$.		

Table 3

Model Summary for Confidence Comparing Gesture and Control Groups

	Coefficient	SE
FIXED-EFFECTS		
(Intercept)	6.56***	0.35
MRatio	0.24	0.15
Group	0.38	0.58
Difficulty	0.03	.27
Sex	0.86	.65
Accuracy	1.08***	.27
Group*Difficulty	-0.47	.51
Group*Sex	0.91	.90
Difficulty*Sex	0.83	.59
Group*Accuracy	0.78 †	.46
Difficulty*Accuracy	-0.97**	.36
Sex*Accuracy	0.79	.54
Group*Difficulty*Sex	-1.10	.82
Group*Difficulty*Accuracy	0.23	.59
Group*Sex*Accuracy	-1.93**	.74
Difficulty*Sex*Accuracy	-0.17	.67
Group*Difficulty*Sex*Accuracy	0.78	.93
	Variance	
RANDOM-EFFECTS		
Intercepts		
Subjects	1.256	
Trials	.02	
<i>Note.</i> Significance codes = *** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .1$.		

Table 4

Model Summary for Accuracy in the Gesture Group

	Coefficient	SE
FIXED-EFFECTS		
(Intercept)	6.56***	.39
MRatio	-0.0006	.22
Gesture Frequency	0.79	.46†
Confidence	1.32***	.29
Difficulty	-0.92*	.36
Sex	-0.30	.56
Gesture Frequency*Confidence	0.32	.48
Gesture Frequency*Difficulty	-0.79	.51
Confidence*Difficulty	-0.85**	.33
Gesture Frequency*Sex	-1.04*	.52
Confidence*Sex	-1.04*	.44
Difficulty*Sex	0.09	.50
Gesture Frequency*Confidence*Difficulty	0.25	.56
Gesture Frequency*Confidence*Sex	0.09	.58
Gesture Frequency*Difficulty*Sex	0.79	.57
Confidence*Difficulty*Sex	0.84†	.50
Gesture Frequency*Confidence*Difficulty*Sex	-0.85	.68
Variance		
RANDOM-EFFECTS		
Intercepts		
Subjects	0.99	
Trials	.00	
<i>Note.</i> Significance codes = ***p < .001, **p < .01, *p < .05, †p < .1.		

Table 5

Model Summary for Confidence in the Gesture Group

	Coefficient	SE
FIXED-EFFECTS		
(Intercept)	6.56***	.39
MRatio	0.04	.22
Gesture Frequency	-0.52†	.46
Accuracy	0.97***	.29
Sex	0.94***	.36
Difficulty	-0.12	.56
Gesture Frequency*Accuracy	0.50	.48
Gesture Frequency*Sex	0.28	.51
Accuracy*Sex	-0.68*	.33
Gesture Frequency*Difficulty	0.58†	.52
Accuracy*Difficulty	-0.44†	.44
Sex*Difficulty	-0.13	.50
Gesture Frequency*Accuracy*Sex	-0.34	.56
Gesture Frequency*Accuracy*Difficulty	-0.32	.58
Gesture Frequency*Sex*Difficulty	-0.31	.57
Accuracy*Sex*Difficulty	0.34	.50
Gesture Frequency*Accuracy*Sex*Difficulty	0.11	.68
Variance		
RANDOM-EFFECTS		
Intercepts		
Subjects	0.13	
Trials	.01	
<i>Note.</i> Significance codes = ***p < .001, **p < .01, *p < .05, †p < .1.		