**How Hand Movement Shapes Communication’s Impact**

**Abstract**

From salespeople and entrepreneurs to politicians and influencers, marketplace actors often communicate with their hands. But despite how integral these movements are to communication, little work in marketing has examined their effects. Might moving one’s hands while speaking increase persuasion? And if so, what types of movements are more impactful? And why? A multimethod investigation, including automated video analysis of thousands of presentations, application of a large multimodal model on almost 200,000 video segments, and preregistered controlled experiments, demonstrates that hand movement can boost impact. Further, the results demonstrate that certain hand gestures (i.e., illustrators) are particularly impactful. By making content easier to understand, using illustrators makes speakers seem more competent, which increases persuasion. Taken together, these findings shed light on hand movement’s impact, deepen understanding around nonverbal communication, and highlight how automated video analysis and multimodal models can provide insights into consumer behavior.

**Keywords**: hand movement, automated video analysis, multimodal models, gestures, nonverbal language, communication

Communicators want to be impactful. Salespeople want to convince customers, entrepreneurs want to persuade investors, leaders want to inspire audiences, and politicians want to garner public support. Accordingly, a burgeoning stream of research has studied how the words or language communicators use shape persuasion, sales, and other outcomes (e.g., Humphreys and Wang 2018; Humphreys, Isaac, and Wang 2021; Moore 2015; Luangrath, Peck, and Barger 2017; Patrick and Hagtvedt 2012a; Wang et al. 2022; Wang, Bendle, and Pan 2024).

But while verbal aspects of communication clearly matter, the non-verbal aspects are also key. When speaking, communicators not only use language, they also “talk” with their hands. Hand movements are an integral part of communication (McNeill 1992), and marketplace actors often move their hands when pitching products, selling ideas, or explaining diagnoses (Clarke, Cornelissen, and Healey 2019; Goldin-Meadow and Alibali 2013; McNeill 1992; Zhou et al. 2021). Indeed, when we analyzed over 2,000 presentations (see Study 2), speakers moved their hands over 80% of the time. Further, such nonverbal communication may have an even bigger impact than the words used (Mehrabian 1971; Michail 2020).

That said, while nonverbal communication is frequent, and important, unfortunately it is relatively understudied in marketing. Further, the little work that has been done has mostly focused on facial expressions (e.g., Zhang et al. 2024), leaving hand movements almost completely unexplored. Might hand movements shape communication’s persuasive impact? And if so, when and how?

Efforts to answer these questions have been stymied by methodological challenges. First, it is difficult to measure hand movements, let alone do so objectively, at scale. This has made it difficult to study how hand movements impact important outcomes in the field. Second, given that hand movements can relate to the words being expressed, truly understanding their effects requires understanding how they connect to the language used.

To address these issues, we use a multi-method approach. Leveraging cutting-edge video processing algorithms (on over 2,100 presentations), and applying a large multimodal model (on almost 200,000 video segments), we identify and classify gestures, enabling scalable, automated measurement of key features in videos. Combining this field data with controlled experiments demonstrates that increased hand movement boosts persuasion (e.g., evaluations and interest in purchase). Further, the results demonstrate that certain hand gestures (i.e., illustrators) are particularly impactful. By making content easier to understand, using illustrators makes speakers seem more competent, which increases persuasion.

This work makes several contributions. First, we shed light on how hand movements shape consumer behavior. While recent work has explored the effects of language (e.g., words or phrases; Humphreys and Wang 2018), and an emerging stream of work is starting to explore vocal features (e.g., pitch or speed; Wang et al. 2021), there has been less attention to body language (Burgoon, Birk, and Pfau 1990). We fill this gap, demonstrating how hand movement shapes communication’s persuasive impact, and underlying processes behind these effects.

Second, we deepen understanding around multimodal communication. While communication often simultaneously occurs through multiple modes (e.g., words, images, and sounds), most existing work has tended to consider individual modes in isolation (e.g., just words, Berger et al. 2020, or just visual content, Hartmann et al. 2021). By analyzing the interplay between words and gestures, we contribute to emerging work (e.g., Ceylan, Diehl, and Proserpio 2024) examining how multiple communication modes *combine* to shape impact.

Third, from a methodological standpoint, we illustrate how automated video analysis can deepen insight into consumer behavior, opening new avenues for future research. Recent work has highlighted the value of video processing for marketing research (Grewal, Gupta, and Hamilton 2021), and begun to apply this approach to online content (e.g., Zhou et al. 2021). But measurement has proven difficult. Video features (e.g., audio, language, and visual elements) interact dynamically over time, making it hard to isolate and measure them in an automated, scalable way (Li, Shi, and Wang 2019). We develop a systematic pipeline for classifying hand gestures in large scale video data (i.e., establishing coding guidelines, training annotators, designing prompts, and refining the model through feedback), providing guidelines on how researchers can use large multimodal models to measure and validate constructs. We also offer a practical tool (available at: <https://chatgpt.com/g/g-QWHBwl1vr-hand-movement-classifier>)[[1]](#footnote-2) to automatically detect and categorize hand gestures in video content.

Finally, these findings have clear practical implications. Communicators are constantly talking with their hands. Our results indicate that simple shifts in the types of movements communicators make can increase their impact. Further, given that 82% of consumers buy products after watching videos (Hoogervorst 2023), understanding how subtle nonverbal cues like gestures can shape persuasion is key.

**Verbal and Nonverbal Communication**

Language shapes almost every marketplace interaction. Indeed, communicators use language to shape perceptions, create connections, and influence others (e.g., Humphreys 2010; Moore 2012; Patrick and Hagtvedt 2012b; see Packard and Berger 2024 for review).

Consistent with its importance, a growing body of research has examined how words shape consumer attitudes and behavior. When service agents refer to themselves as “I” rather than “we,” for example, it improves customer satisfaction because it makes customers feel like the agent cares and is involved (Packard, Moore, and McFerran 2018). Similarly, swear words make reviews more helpful because they convey meaning about the reviewer and the product (Lafreniere, Moore, and Fisher 2022) and refusing things by saying “I don’t” rather than “I can’t” can enhance persuasion (Patrick and Hagtvedt 2012a).

But while verbal communicationhas attracted lots of recent interest, there’s been less attention to *nonverbal* communication, or how marketplace actors communicate with their voice or body. Some work has begun to explore vocal features, finding that lower pitched voices can lead consumers to infer advertised products are larger (Lowe and Haws 2017) or that certain vocal tones (i.e., those denoting focus or stable emotions) can increase persuasion (Wang et al. 2021). Other work has started to explore things like posture and facial expressions, finding that erect postures make speakers seem more powerful (Briñol et al. 2017), and that smiling makes communicators seem more sociable and warmer (e.g., Wang et al. 2017; Zhang et al. 2024).

That said, while this emerging work has provided valuable insights, it has left another key type of nonverbal communication relatively understudied: hand movements.

**Hand Movement**

Speech is often accompanied by hand movement. Whether explaining things to themselves or others, communicators frequently move their hands (e.g., Burgoon, Manusov, and Guerrero 2021; Goldin-Meadow and Beilock 2010). Further, such movements help communicators organize their thoughts and articulate ideas (Goldin-Meadow 1999). When someone is trying to recall their shopping list, for example, they might count on their fingers to remember what to buy. Indeed, moving one’s hands can help people’s thinking and learning across various tasks, including word learning (McGregor et al. 2009), problem solving (Beilock and Goldin-Meadow 2010), and math (Alibali and DiRusso 1999).

But while it’s clear that hand movements can help the people that use them, might they also impact observers? And if so, how? Take a salesperson pitching a product or an entrepreneur pitching an idea. Might moving their hands more while speaking make them more persuasive?

While we are not aware of work that examines these questions, some work in related disciplines has begun to explore other impacts of hand movements. Research on deception, for example, has examined the link between hand movements and lying. People who are telling the truth tend to point to physically present objects (Caso et al. 2006), for example, and individuals who exhibit more different movements tend to be more honest (Burgoon, Schuetzler, and Wilson 2015). Similarly, studies find that observers think communicators who place their hand on their heart are more honest (Parzuchowski and Wojciszke 2014) and infer more about whether others are lying from their body movements than their facial expressions (Ekman and Friesan 1974).

Research in education, communication, and psychology has also examined how hand movements might influence memory (see Hotstetter 2011 and Dargue, Sweller, and Jones 2019 for reviews). Compared to just hearing words, for example, seeing speakers move their hands can help observers recall how to get somewhere (Austin, Sweller, and Van Bergen 2018), remember math formulas (Cook et al. 2017), and memorize foreign language (Gluhareva and Prieto 2017).

But while scholars in other disciplines have provided insight into how hand movements relate to deception or memory, key gaps remain. First, to the best of our knowledge, no marketing papers have examined how hand movement influences consumer responses. Given that entrepreneurs, salespeople, influencers, and other marketplace actors “talk” with their hands, understanding how hand movement might shape communications’ persuasive impact seems critical. Second, given prior work (in other fields) has compared adding hand movements to speech alone (Dargue, Sweller, and Jones 2019), it remains unclear whether specific *types* of movements are more effective.[[2]](#footnote-3) Third, studies have mostly relied on manual coding, which is labor-intensive, costly, and difficult to scale (c.f. Burgoon, et al. 2015).

**Hand Movement and Communication’s Persuasive Impact**

We suggest that hand movements should boost the persuasive impact of communication. While this prediction has never been empirically tested, building on prior work, we suggest several reasons this could occur.

Hand movements could increase *understanding*. Visually representing spoken content can make it more concrete, helping to clarify communication’s meaning (e.g., McNeill 1992; Lurie and Mason 2007). When someone says, “the structure of our business model is layered,” for example, and uses their hands to show layers stacked on top of each other, it gives listeners a better sense of what the speaker means (Paivio 1969). Indeed, hand movements can disambiguate speech’s meaning (Kelly et al. 1999) and increase learning (Ping and Goldin-Meadow 2008).

By making things more understandable, hand movements could, in turn, make speakers seem more *competent*. It’s hard to explain something well if you don’t understand it, so being able to explain something in an understandable way (e.g., representing a layered structure) should signal greater knowledge and mastery over a topic. It suggests that the speaker has internalized the structure and grasped the underlying logic deeply enough to convey it clearly. Consequently, to the degree hand movements increase understanding, they may also increase perceived competence. Note that this may be particularly true in multi-modal communication (i.e., moving hands while speaking) because of the effort and expertise required. Being able to figure out (and make) the right movements to deepen understanding of what is being said indicates skill and ability. Indeed, communicators who are able to draw visual aids (e.g., diagrams or charts) to explain things are seen as more knowledgeable (Chi, Feltovich, and Glaser 1981) and confident with the subject matter (Garcia-Retamero and Cokely 2017). Seeming more competent, in turn, should lead communicators to be evaluated more positively and increase persuasion (e.g., Dubois, Rucker, and Galinky 2016; Fiske, Cuddy and Glick 2007).

Hand movements might also encourage audiences to pay *attention*. Movement naturally draws attention (Franconieri and Simons 2003), and pointing with a pen or moving objects within a visual scene guides people’s focus to critical information (Abrams and Christ 2003). Hand movements might have similar effects, drawing the audience in and encouraging them to attend to what is being said (Pi, Hong, and Yang 2017). Given enhanced attention may increase involvement (Mrvka, Westfall, and Van Boven 2009), this could encourage persuasion (e.g., Kang and Tversky 2016).

Finally, hand movement might make speakers seem more *extraverted* or *emotionally involved.* Extraverted people move their hands more (Carlson et al. 2016), and hand movements can signal emotional involvement (Asalıoğlu and Göksun 2023). Excited people jump (Mortillaro and Dukes 2018), for example, and proud people raise their fists as a sign of victory (Dael, Mortillaro, and Scherer 2012). Perceiving speakers as more emotionally involved could lead audiences to respond more positively (e.g., Rocklage, Rucker, and Nordgren 2018).

Taken together, these different aspects should lead hand movements to encourage positive evaluations, both of the speaker and the thing being pitched. Study 1 begins to explore this possibility. Data, code, and materials for all studies are available at <https://osf.io/yjez7/?view_only=e5d2296924c0417d91ba383f98d74750>.

**Study 1: Hand Movement in the Field**

To begin to test hand movement’s potential impact, Study 1 turns to the field. Using automated video analysis, we measure hand movement in thousands of presentations, testing whether (controlling for various other factors) presentations are evaluated more positively when speakers move their hands more.

**Method**

We obtained data on TED Talks uploaded to the TED.com website from June 27, 2006, to September 21, 2017 (<https://www.kaggle.com/datasets/rounakbanik/ted-talks>). We focused on TED Talks for several reasons. First, they follow a consistent format (i.e., speakers standing on a stage, with consistent camera framing), enabling more precise measurement of movement. Second, they provide a measure of consumer response (e.g., how many viewers like each video). Third, they involve a diverse set of speakers, talking about a range of topics, allowing us to test the generalizability of the effect.

For each talk, the dataset contains information useful to form control variables, including the gender and occupation of the speaker, a transcript of the talk, the date it was recorded, its length, the video category (e.g., comedy, education, and politics), and number of languages for subtitles. Given the video recordings themselves are not included, we scraped all available ones from YouTube. Ignoring videos featuring multiple speakers or focused on performance (e.g., music or magic tricks where hands are used for things other than speaking) left a final sample of 2,184 videos from 1,857 speakers.

Automated video analysis was used to quantify hand movement. First, each video was divided into frames using the Open-Source Computer Vision (OpenCV) Library. The frame rate of a video is determined based on the metadata stored within the video file itself.

Second, for each frame, we detected key points of each hand and computed the spatial distance (or movement) between consecutive frames.[[3]](#footnote-4) Following recent advancements in visual processing (e.g., Vakunov et al. 2020; Remiro, Gil-Martín, and San-Segundo 2023), we used Google’s MediaPipe Hands module. This module detects hands in images and localizes 21 landmarks or key points on the hand, such as fingertips, knuckles, and palm (see Figure 1).[[4]](#footnote-5) The model was trained on approximately 30,000 real-world images, as well as several rendered synthetic hand models imposed over various backgrounds, achieving an average precision of 95.7% in landmark detection.

**Figure 1**: Hand Landmark Detected by Google’s MediaPipe

A screenshot of a computer

Description automatically generated

To quantify dynamic changes in hands’ position, we used the Euclidean distance:

|  |  |
| --- | --- |
|  | **(1)** |

where the vector (x2*i*, y2*i*) represents the X and Y coordinates of key points in the current frame while the vector (x1*i*, y1*i*) represents the X and Y coordinates of the same key points in the previous frame (see Figure 2 for example of sequential video frames).

**Figure 2**: Example of Hands Movement Detection

|  |  |  |
| --- | --- | --- |
| A person in a suit with microphones  Description automatically generated | A person in a suit  Description automatically generated | A person in a suit  Description automatically generated |

Third, to account for changes in hand size (e.g., due to variations in camera distance or zoom settings), we used a scale factor to normalize the movement, ensuring more accurate measurement (see “Movement Normalization” in Web Appendix A).

Fourth, we computed average movement per frame. This was calculated as the total movement divided by the number of frames in which hands were detected (M = .50, SD = .27). This measure correlates well with human perceptions (*r* = .74), underscoring its validity.[[5]](#footnote-6)

Fifth, to capture consumer response, following prior work (e.g., Chung, Ding, and Kalra 2023, Valsesia, Proserpio, and Nunes 2021), we collected the number of likes each video received on YouTube. We examined number of likes for several reasons. First, it is a good measure of how much the content resonated with the audience (Cascio Rizzo et al. 2023, 2024). Second, a great deal of research finds that such engagement is correlated with sales (e.g., Beichert et al. 2024; Kumar et al. 2016). Third, from a practical perspective, it serves as a simple, straightforward measure of content effectiveness that content creators use to drive communication strategy (Liadeli, Sotgiu, and Verlegh 2022). Given the number ranges from 40 to 1,873,539, and is skewed (skewness = 14.22, SE = .001), it was log-transformed.

Finally, ordinary least squares regression examined the relationship between hand movement and likes. All continuous predictor variables were standardized (*z*-scored).

**Results**

As predicted, presentations were evaluated more positively when speakers moved their hands more (*b* = .262, SE = .034, *t* = 7.69, *p* < .001; Table 2, Model 1).

*Control Variables.* While this initial result is intriguing, one could wonder whether it is driven by other factors. Consequently, to account for unobserved heterogeneity, we control for a range of factors (see Table 1 for a full list of the measures, their operationalizations, and rationale for the controls, Table WA1 for summary statistics, and the Web Appendix A for further details). First, results might be driven by aspects of the speaker, so we control for their profession and demographics (e.g., gender and age). Second, the content speakers deliver also likely plays a role, so we control for aspects of what speakers say (e.g., topics discussed, emotionality, and concreteness), how they say it (e.g., speaking rate and pitch variation), and their facial expressions. Third, aspects of the video may also affect evaluations, so we control for video duration, motion area, and visual aesthetics (e.g., saturation and clarity). Finally, we control for aspects of the talk itself, such as number of video views and time effects.

The full model takes the following form:

|  |  |
| --- | --- |
| Yi = β1 Hand Movementi + **X**′i γ + εi, | **(2)** |

where the dependent variable Yi is the (log) number of likes for video *i*. The focal independent variable Hand Movementi indicates the average hand movement in the presentation *i*. The matrix **X**′i is the set of covariates.

*Results Including Controls*. Even after accounting for all these controls, however, presentations were still evaluated more positively when speakers moved their hands more (*b* = .028, SE = .008, *t* = 3.65, *p* < .001; Table 2, Model 2). Results indicate that doubling the amount of movement, for example, is associated with an average 5.18% increase in (or 798 additional) likes.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Operationalization** | **Rationale and Related Studies** |
| Occupation | Speaker’s occupation (see Table WA2) | Certain professions may be evaluated more positively (Chaiken and Maheswaran 1994) |
| Gender | Man or woman (dummy) | Certain genders may be evaluated more positively (Fiske 2017) |
| Age | Estimated age using Face++ | Age may influence speaker evaluations (Fiske 2017) |
| Skin Tone | Skin darkness using Face++ | A speaker’s race may affect evaluations (Fiske 2017) |
| Topics | Proportion of LDA topics | Certain themes or topics may be received more positively (Berger et al. 2020) |
| Emotionality | Emotionality measure from the Evaluative Lexicon | Emotional language may complement hand movements, affecting evaluations (Rocklage, Rucker, and Nordgren 2018) |
| Concreteness | Concreteness ratings, Paetzold and Specia (2016) | Concrete language may complement hand movements, boosting evaluations (Packard and Berger 2021) |
| Narrativity | LIWC Narrativity dimension | Hands may help tell a story, which increases liking (Escalas 2004) |
| Focus Past | LIWC Focus Past dimension | Hands may be used to discuss past events, affecting evaluations (Packard, Berger, and Boghrati 2024) |
| Questions | LIWC Questions dimension | Questions may enhance liking (Huang et al. 2017) |
| Positivity | LIWC Tone dimension | Positive language may complement hand movements, affecting evaluations (Berger et al. 2020) |
| Word Familiarity | Familiarity ratings, Paetzold and Specia (2016) | Familiar words may be easier to process, increasing liking (Pancer et al. 2019) |
| Swear Words | LIWC Swear dimension | Swear words may convey meaning, increasing persuasion (Lafreniere, Moore, and Fisher 2022) |
| Function Words | LIWC Function Words dimension | Function words reveal psychological states, which affect persuasion (Sela, Wheeler, and Sarial-Abi 2012) |
| LIWC Main Dimension, LIWC Psych Processes | Analytic, Clout, Authenticity, and Cognition, Drives, Social, Time Orientation, Conversation | Psychological states of the speaker can affect evaluations (Packard and Berger 2021) |
| Speaking Rate | Average words per minute | Faster speakers may seem more confident, increasing persuasion (Cesario and Higgins 2008) |
| Vocal Pitch, Pitch SD | Fundamental frequency (Hz) using Praat | Lower pitch may convey confidence, which increases persuasion (Van Zant and Berger 2020) |
| Vocal Volume, Volume SD | Sound intensity level (dB) using Praat | Louder speakers may seem more confident, increasing persuasion (Van Zant and Berger 2020) |
| HNR  Shimmer  Facial Emotional States | Proportion of harmonic sound energy to noise energy using Praat  Cycle-to-cycle variation in the amplitude of the vocal signal using Praat  Anger, disgust, fear, happiness, neutral, sadness, surprise using Face++ | Harmonics-to-Noise Ratio (HNR) is linked to confidence (Hildebrand et al. 2020)  Local shimmer is linked to confidence (Hildebrand et al. 2020)  Facial emotions may influence evaluations (Li and Xie 2020) |
| Duration | Video length in minutes | Longer videos may include more movement, affecting evaluations |
| Motion Area, Magnitude, Direction | Area (in pixels), magnitude (in pixels), direction (in degrees) using OpenCV algorithm | Hand framing may influence evaluations (Zhou et al. 2021) |
| Average Scene Length | Total scene duration/number of scenes using PySceneDetect | Changes in framing may make videos more enjoyable (Zhou et al. 2021) |
| WHP, Saturation, Brightness, Contrast of Brightness, Clarity | Pixel-level color features using OpenCV algorithm | Visual aesthetic features may influence emotions and attention, and so evaluations (Zhou et al. 2021) |
| Views | Log number of video views | Videos with more views may receive more likes |
| Language Subtitles | Number of available subtitles | Multiple subtitles may increase reach and thus likes |
| Category | Video category (see Table WB3) | Broad-interest categories may be more liked |
| Time Fixed Effects | Year, month, day fixed effects | Certain times of year or specific annual events may get more likes |
| Time Difference | Days between publishing and data scraping | Older videos may accumulate more likes |

**Table 1**: Control Variables and Operationalizations

**Table 2**: Movement and Consumer Response

|  |  |  |  |
| --- | --- | --- | --- |
|  | **(1)** | **(2)**  **Including Controls** | **(3)**  **Movement Type** |
| **Hand Movement**  **Illustrators**  **Highlighters**  **Unrelated**  **No Movement**  Controls  *Speaker*  Occupation  Gender  Age  Skin Tone  *Content*  Topics  Emotionality  Concreteness  Narrativity  Focus Past  Questions  Positivity  Word Familiarity  Swear Words  Function Words  LIWC Main Dimensions  LIWC Psych. Processes  Speaking Rate  Vocal Pitch  Vocal Pitch SD  Vocal Volume  Vocal Volume SD  HNR  Vocal Shimmer  Facial Emotional States  *Video*  Duration  Motion Area  Motion Magnitude  Motion Direction  Average Scene Length  Warm Hue Proportion  Saturation  Brightness  Contrast of Brightness  Clarity  *Talk*  Views  Language Subtitles  Category  Time Fixed Effects  Time Difference | **.262\*\*\*(.034)** | **.028\*\*\*(.008)**  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included | –  **.023\*\*\*(.010)**  **.008\*\*\*(.012)**  **.002\*\*\*(.014)**  –**.013\*\*\*(.011)**  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included  Included |
| N | 2,184 | 2,184 | 2,184 |

*Notes*: \*\*\**p* < .001, \*\**p* < .01, \**p* < .05. Full results for controls are in Table WB3, Model 1.

**Robustness**

To test robustness to a different outcome measure, we scraped 446,886 YouTube consumers’ comments and used Evaluative Lexicon’s “valence” dimension (Rocklage, Rucker, and Nordgren 2018) to measure responses to the talk. Results remain the same. Responses were more positive when speakers moved their hands more (*b* = .069, SE = .018, *t* = 3.94, *p* < .001; Table WB4, Model 2).[[6]](#footnote-7)

Additional analyses test the robustness of our measure, model specification, and other potential alternative explanations. See Table 3 for an overview, Web Appendix B for details on each test, and Table WB4 for results.

**Table 3**: Robustness Checks

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | |  | **What We Test** | |  | | **How We Test It** | |
| **Measurement** | | | |  |  | |  | |  | |
|  | *Alternate measure of movement* |  | | | | Could the effect be driven by the  way movement is measured? |  | Results hold using total movement (*b* = .018, SE = .009, *t* = 2.01, *p* = .044; Table WB3, Model 3) or an analysis every 100 frames (*b* = .018, SE = .008, *t* = 2.28, *p* = .023; Table WB3, Model 4). | | |
| **Modeling** | | |  |  | | |  |  | |
|  | *Alternate*  *approach*  *Alternate DV*  *Autocorrelation* |  | | | | Are the results driven by the modeling approach used?  Could the effect be driven by the dependent variable used?  Could the effect be driven by the number of likes within a speaker being correlated across talks? |  | Results hold using Lasso penalized linear regression (*b* = .027, SE = .007, *t* = 3.66, *p* < .001; Table WB3, Model 5) or negative binomial regression (*b* = .030, SE = .007, *t* = 4.22, *p* < .001; Table WB3, Model 6).  Results hold using liking rate (i.e., likes number/views number; *b* = .0005, SE = .0001, *t* = 5.10, *p* < .001; Table WB3, Model 7).  Results hold including cluster standard errors at the speaker level (*b* = .028, SE = .008, *t* = 3.57, *p* < .001; Table WB3, Model 8). | | |
| **Alternative Explanations** | | | | | | | | | |
|  | *Talk Title* |  | | | | Could the effect be driven by talk titles? |  | Results hold controlling for talk title topics (*b* = .029, SE = .008, *t* = 3.68, *p* < .001). | | |
|  | *Within-presentation movement variability* |  | | | | Could the effect be driven by variation in speakers’ movement within the presentation? |  | Variation in average speakers’ movement is relatively low (i.e., .22), and results hold controlling for movement SD (*b* = .025, SE = .008, *t* = 3.11, *p* = .002). | | |

**Discussion**

Overall, Study 1 provides preliminary evidence regarding hand movement and observer responses. Automated video analysis of thousands of presentations demonstrates that presentations are evaluated more positively (i.e., receive more likes and positive responses) when speakers move their hands more. Finding similar results including dozens of controls, and testing various measurement and model specifications, casts doubt on alternative explanations.

*Quadratic Effect?* One could wonder whether hand movement has a quadratic effect. Maybe increased movement is beneficial up to a certain point (e.g., the average amount of movement) but moving hands too much could backfire. A quadratic term is not significant (*b* = .001, SE = .004, *t* = .04, *p* = .682), though, casting doubt on this possibility.[[7]](#footnote-8)

**Which Movements Are More Beneficial?**

The results of Study 1 are intriguing, but are *all* hand movements beneficial? Or might certain movements be more beneficial than others? To address these questions, we go beyond merely asking *whether* hand movement is beneficial to ask *which* hand movements might have more impact, and *why*.

**Types of Hand Movement**

Some have speculated that hand movements with a specific communication goal (i.e., gestures, or movements that carry meaning which complements or supplements speech; Ekman and Friesen 1972) should be more effective (e.g., Kendon 2004). Other theoretical work has attempted to classify gestures based on their communicative function (McNeill 1992), categorizing them as either illustrating or highlighting something.

Building on McNeill’s (1992) framework, and subsequent meta-analytic work (Dargue, Sweller, and Jones 2019), we examine two types of gestures (see Table 4 for how our taxonomy relates to prior work, and Table 5 for examples).

**Table 4**: Gestures Taxonomy

|  |  |  |  |
| --- | --- | --- | --- |
| **McNeill’s (1992) Classification** | | **Our Classification** | |
| Gesture Type | Description | Gesture Type | Description |
| ***Iconic***  ***Metaphoric*** | Visually represents concrete action, event, or object that is verbally described.  Visually represents an abstract concept or a metaphor rather than a concrete object or event. | *Illustrator* | Visually represents concrete objects, actions, events, abstract concepts, or metaphorical expressions. |
| ***Deictic***  ***Beat*** | Indicates a concrete object, direction, or event through pointing.  Emphasizes or accentuates speech. They are typically rhythmic, flicking movements of the hands that accompany speech. | *Highlighter* | Emphasizes spoken content (without visually depicting things) or draws attention to something physically present or. |

Some gestures *illustrate* what is being talked about, and the form they take is semantically related to the content of the accompanying speech (Dargue and Sweller 2018). Making an upward line in the air, for example, while saying “the sales have increased steadily” or tracing the steps of stairs while saying “the steps of success” both use hand movements to create a visual image of what is being discussed. These types of gestures can be described as *illustrators* (Ekman and Friesen 1972; McNeill 1992).

Other gestures *highlight* or draw attention to what was said, without illustrating it. Placing one’s hands on their head while saying “Oh My God,” for example, or raising a bottle while saying “this water bottle,” serves to emphasize the spoken content or draw attention to something physically present.[[8]](#footnote-9) We describe these types of gestures as *highlighters.*

**Table 5**: Examples of Illustrators, Highlighters, and Unrelated Movements

|  |  |  |  |
| --- | --- | --- | --- |
| **Movement Type** | **Speech** | **Movement Description** | **Reasoning** |
| **Illustrator** | “I was drinking water” | Drinking gesture | Visually depicts the act of drinking. |
| “I caught a fish this big” | Extending hands apart to show the size of the fish | Visually depicts the size of a concrete object. |
| “I was trying to remember it” | Mimicking the turning of gears | Visually depicts the process of retrieving information. |
|  | “These are the best shoes” | Pointing to the shoes | Draws attention to something physically present. |
| **Highlighter** | “This is the best thing ever!” | Fists up (in joy) | Emphasizes the emotional content of the speech. |
|  | “It wasn’t me” | Both hands up facing the audience | Emphasizes the speaker’s innocence without visually representing it. |
| **Unrelated** | “It was a great weekend”  “We should go with option B” | Scratching an itch  Adjusting glasses | Unrelated to the spoken content and serves no communicative purpose.  Unrelated to the spoken content and serves no communicative purpose. |
| “That is an interesting project” | Opening and closing hands randomly | Unrelated to the spoken content and serves no communicative purpose. |

Note, this two-class categorization aligns with McNeill (1992), but groups iconic and metaphorical gestures into “illustrators,” and deictic and beat gestures into “highlighters” (Table 4). This was done, in part, to enable the LMM used in Study 2 to effectively analyze and categorize gestures accurately and consistently. That said, to ensure that the effects are not driven by this simpler classification, Study 3b uses both iconic and metaphorical gestures as illustrators, and deictic and beats gestures as highlighters.

In addition, beyond examining illustrators and highlighters, we also examine hand movements which have no communicative goal. When speakers scratch an itch, for example, it is usually unrelated to what is being said and is not intended to communicate meaning.[[9]](#footnote-10) Similarly, speakers might stretch their hands or make superfluous movements while talking. We describe these as *unrelated movements*.[[10]](#footnote-11)

Importantly, while prior work has examined how gestures impact memory (see Hofstetter 2011; Dargue, Sweller, and Jones 2019 for reviews), no work has examined how they might impact more marketing related outcomes. Further, prior work found inconsistent results, making it hard to know whether certain types of gestures are more impactful. While Beattie and Shovelton (1999) found that illustrators enhanced memory, for example, and Cook, Duffy, and Fenn (2013) found that highlighters enhanced memory, other studies found no effect for either (Gluhareva, and Prieto 2017). Along these lines, meta-analytic work on memory found no effect of gesture type (Dargue, Sweller, and Jones 2019),[[11]](#footnote-12) so it remains unclear whether certain types of gestures are more impactful than others in shaping consumer behavior.

**The Impact of Illustrators**

Building on some of the potential mechanisms outlined for why hand movements might be beneficial, we suggest that a particular type of gesture (i.e., illustrators) may have greater persuasive impact. Unlike other movements, illustrators provide a visual representation of the verbal message (Burgoon, Birk, and Pfau 1990; Goldin-Meadow 1999; Goldin-Meadow and Alibali 2013; McNeill 1992), which should make content easier to understand, and speakers seem more competent. Mimicking the circular motion of gears while saying “the gears rotate this way,” for example, creates a visual image of the spoken content, which should deepen understanding.

Indeed, matching content across multiple formats (e.g., text and visuals) can enhance fluency (Winkielman et al. 2012). Ceylan, Diehl, and Proserpio (2024), for example, found that greater similarity between written text and images (e.g., a review about a hamburger paired with a picture of it) makes reviews more helpful because it makes them easier to process. Similarly, when saying “the Earth is round,” speakers could visually depict the shape of the Earth with their hands (illustrators) rather than raising an index finger to punctuate the statement (highlighter), waving hands without intent (unrelated movement), or remaining still (no movement). By providing the same information through multiple modes (i.e., words and visual gestures), illustrators’ semantic integration of speech and gesture should make the content easier to process, enhancing understandability (Carney and Levin 2002; Cook 2006; Dargue and Sweller 2018).

Increasing understanding, in turn, can make speakers seem more competent. Representing the Earth’s shape not only demonstrates understanding of the content, but also the ability to effectively convey it through visual means. After all, translating concepts into accessible, tangible forms requires both fluency with, and knowledge of, the subject matter (Burgoon, Buller, and Woodall 1996). Consistent with this notion, drawing diagrams or maps to visually represent spoken content can make communicators seem more knowledgeable (Chi, Feltovich, and Glaser 1981; Garcia-Retamero and Cokely 2017). Similar effects may occur for illustrators. By making ideas easier to understand, illustrators should signal that the speaker is more knowledgeable about, and confident with, the subject matter.[[12]](#footnote-13)

This, in turn, should increase communication’s persuasive impact (see Figure 3 for the full conceptual model). By making speakers seem more competent, using illustrators should lead them (and their communications) to be evaluated more positively, and increase the likelihood that those communications drive action (Fiske, Cuddy, and Glick 2007; Wang et al. 2017).

**Figure 3**: Conceptual Model

A black and white photo of words

Description automatically generated with medium confidence

Note, we are not claiming that the other mechanisms mentioned earlier do not play any role. Illustrators may attract more attention than unrelated movements, for example, and signal more emotional involvement than no movement at all. That said, it is less clear why illustrators would capture more attention than highlighters, or signal more emotional involvement than unrelated movements (Asalıoğlu and Göksun 2023). We examine potential underlying mechanisms more directly in Studies 3a and 3b.

**Study 2: Impact of Illustrators**

Study 2 uses a large multimodal model to detect the use of illustrators, highlighters, unrelated movements, and no movement in almost 200,000 video segments from the talks from Study 1, testing whether presentations are evaluated more positively when speakers use illustrators.

Testing this, though, requires doing more than just looking at movements in isolation. After all, the same movement could be an illustrator or unrelated, depending on the language it is being produced with. Consequently, truly exploring the impact of hand movements requires understanding their potential connection with the language used. To do so, Study 2 explores the relationship between movements and concurrent language and uses that to test the relationship between the different types of movements and consumer response.

**Method**

To automatically detect different types of hand movement, we used Google’s Gemini 1.5 Pro, the most advanced Gemini version available when the analysis was performed. Gemini is a Large Multimodal Model (LMM) that integrates multiple modes of information (e.g., text, audio, and visual) to perform complex tasks that require understanding the relationships between these various forms of input. It has been shown to perform better than other models on many different tasks (Google DeepMind 2024). We also explored other LMMs such as GPT-4o and LLaVA-NeXT, but Gemini performed better on our task of hand movement classification.

The analysis followed seven main steps (see Table 6 for an overview). First, we established clear coding guidelines in our prompt, including definitions of movement types, examples, and rules for coding (see “Instructions for Gemini” section in the Web Appendix A).

Second, to evaluate Gemini’s performance, manual coding was used. Seven research assistants went through four weeks of training (to ensure reliability) and annotated 1,518 gestures from 130 videos.

**Table 6**: Measuring and Validating Constructs through Gemini

|  |  |
| --- | --- |
| **Step** | **Description** |
| 1 | Define the key construct(s) and set up the rules for coding |
| 2 | Train human annotators and generate an annotated sample |
| 3 | Prompt the LMM based on the guidelines and few-shot examples from steps 1-2 |
| 4 | Incorporate a feedback loop to refine coding instructions and assess LMM’s accuracy |
| 5 | Test and assess LMM’s performance on a different sample |
| 6 | Deploy the LMM to perform the coding task on the whole sample |
| 7 | Review and assess the annotation accuracy through human raters |

Third, Gemini was prompted using the coding instructions and few-shot examples from the annotated sample. We also began evaluating Gemini’s performance by classifying hand movement in videos that had already been coded by humans.

Fourth, multiple rounds of iterative feedback were conducted to refine coding rules and improve classification accuracy (see “Solving LMM’S Classification Ambiguities” section in the Web Appendix A). Once classification accuracy reached a reasonably high level (accuracy = .803, precision = .841, recall = .803, F1 score = .813) we proceeded to the next step.

Fifth, Gemini’s performance was tested on a different sample of annotated data, and accuracy was assessed.

Sixth, Gemini was asked to classify hand movement types in all videos. In particular, each video was divided into 10-second segments, generating 191,548 snippets. Given the multimodal nature of the task, which requires understanding how hand movements relate to what is being said, each video snippet included the real-time transcript so that Gemini could evaluate the movement in relation to the spoken content.[[13]](#footnote-14) To the best of our knowledge, this represents a novel application of large multimodal models (LMMs) for classifying naturalistic gesture types at scale in behavioral research. At the end of this phase, Gemini classified video segments as including illustrators (M = 5.84 at the talk level, or 6.5% of total movements), highlighters (M = 21.37), unrelated movements (M = 45.33), or no movement (M = 14.53). For examples of illustrators, highlighters, and unrelated movements as detected by Gemini, see Figure 4.

**Figure 4**: Example of LMM’s detection of Gestures

**Gemini’s Classification: Illustrator**

|  |
| --- |
| A person in a black shirt  Description automatically generated A person in a black shirt  Description automatically generated A person in a suit  Description automatically generated |
| **Speech:** “Instead of having paper in my chart, and it being this thick”.  **Gemini’s reasoning:** The speaker is using his hands to visually depict the thickness of the paper in a chart. Although he mentions “my chart” and “your chart”, neither of them is physically present, making this gesture an Illustrator as it depicts a concrete object (paper’s thickness) that is not physically there.  **Gemini’s Classification: Highlighter** |
| A person holding a bottle of water  Description automatically generated A person holding a bottle of water  Description automatically generated A person holding a bottle of water  Description automatically generated |
| **Speech:** “So, that pathologist is just like looking at this plastic bottle”.  **Gemini’s reasoning:** The speaker is holding up a plastic bottle while saying “this plastic bottle”. Since the object being referred to is physically present and he's drawing attention to it, the gesture is classified as a Highlighter.  **Gemini’s Classification: Unrelated Gesture** |
| A person standing in front of a blue background  Description automatically generated A person holding a microphone  Description automatically generated A person speaking into a microphone  Description automatically generated |
| **Speech:** “We can prevent most cancers”.  **Gemini’s reasoning:** The speaker is holding a clicker, and his hand movements seem to be incidental adjustments related to holding the clicker rather than deliberate gestures. There's no clear connection between these movements and the spoken content about preventing cancer. Therefore, it's classified as an Unrelated Gesture. |

Seventh, to assess classification accuracy, two trained research assistants (intercoder agreement Cohen’s κ = .74) manually coded a random sample of 200 video snippets. They agreed that Gemini committed errors in 18% of cases, resulting in an accuracy of 82%.

Next, we examined the link between types of hand movement and consumer responses. The model takes the following form:

|  |  |
| --- | --- |
| Yi = β1 Illustratorsi + β2 Highlightersi + β3 Unrelatedi + β4 No Movementi + **X**′i γ + εi, | **(3)** |

where Illustratorsi, Highlightersi, Unrelatedi, No Movementi indicate the number of illustrators, highlighters, unrelated movements, and no movement in the presentation *i*, respectively. The dependent variable, controls, and analysis approach are the same as in Study 1.

**Results**

Results indicate that audiences responded more favorably when speakers used more illustrators (*b* = .023, SE = .010, *t* = 2.44, *p* = .015). Highlighters (*b* = .008, SE = .012, *t* = .66, *p* = .512), unrelated movements (*b* = .002, SE = .014, *t* = .14, *p* = .885) and no movement (*b* = –.013, SE = .011, *t* = –1.16, *p* = .247) did not have the same positive effect (see Table 2, Model 3).

**Robustness**

One could wonder if the results are somehow driven by the automated tool used to classify movements (e.g., LMM miscalibration). Consequently, to test robustness, we had two trained research assistants code all gestures in 130 videos (see Web Appendix A and Table WA6 for details). Results remain similar. While illustrators were associated with more favorable consumer response (*b* = .290, SE = .094, *t* = 3.07, *p* = .006), highlighters (*b* = –.010, SE = .138, *t* = –.07, *p* = .945), unrelated movements (*b* = .375, SE = .389, *t* = .97, *p* = .342), and no movement (*b* = .157, SE = .180, *t* = .87, *p* = .393) didn’t have any impact.

One might also wonder whether the way movement types were operationalized drove the effect. Using the proportion of movement types rather than the count, though, yields the same results.[[14]](#footnote-15)

**Discussion**

Results of Study 2 provide deeper insight into potential effects of hand movements. Analyzing almost 200,000 video segments demonstrates that not all movements are beneficial. While audiences responded more favorably to presentations that used illustrators, other movements (e.g., highlighters and unrelated movements) had no effect.

**Study 3a: Experimentally Manipulating Types of Hand Movement**

While results of the first two studies shed light on the potential impact of hand movements, one could wonder whether the relationship is truly causal. Controls cast doubt on alternative explanations, but an even stronger test would be to experimentally manipulate types of hand movement and measure their impact. Study 3a (preregistered: <https://aspredicted.org/c92n-x6r4.pdf>) does this, testing whether illustrators have the most positive impact.

Second, the study tests generalizability. We use a different context (an entrepreneur’s pitch) and assess whether the results extend to less skilled communicators, using a normal undergraduate research assistant as the speaker. In addition, we examine whether the effects extend to other dependent variables. Beyond how their presentation is evaluated, marketplace actors also care what the audience thinks about the product, service, or brand they are discussing. Consequently, in addition to evaluations of the speaker, we also examine whether illustrators make audiences like the product more and increase their interest in purchasing it.

Third, we begin to explore a potential underlying process. We have suggested that illustrators make content easier to understand, which makes the speaker seem more competent, which boosts the persuasive impact of their communication. Study 3a tests this process.

Finally, we test an alternative explanation based on movement typicality (i.e., whether illustrators are simply more typical).

**Method**

Participants (N = 685, Prolific) watched a short video of an entrepreneur pitching a new skincare product. They were randomly assigned to a condition in a 4 (movement type: illustrator vs. highlighter vs. unrelated vs. no movement) between-subjects design.[[15]](#footnote-16)

To ensure experimental control, the “entrepreneur” was a trained research assistant who went through intensive training to ensure conditions were equivalent on all other dimensions (see below for tests). See Web Appendix B for exclusions, demographic, and experimental materials. The only difference between conditions was the type of hand movement (the words were identical). The speaker either used no hand movements, or used one, brief gesture that either illustrated, highlighted, or was unrelated to what they were talking about. While saying, “just spread it on your face,” for example, the speaker either made a circular motion on the face (Illustrator), pointed to the face (Highlighter), moved their hands without any specific communication goal (Unrelated), or didn’t move their hands (No Movement). The research assistant was instructed to keep their voice (e.g., vocal pitch, intonation, loudness, pausing, and emphasis), facial expressions, movement outside of the gesture, and recording duration the same across recordings. Ancillary analyses show that they did so.[[16]](#footnote-17)

After watching the video, participants completed the dependent variables. They were asked how much they liked the speaker (1 = don’t like them at all, 7 = like them a great deal), how much they liked the product (1 = don’t like it at all, 7 = like it a great deal), and their interest in buying it (“Imagine you are looking for a product like this. How interested would you be in buying this particular one?” 1 = not at all interested, 7 = extremely interested). Given that principal components analysis found that three items all form a single component (all loadings > .88), following the pre-registration, they were averaged to form an index of persuasion (α = .86), but results are the same when each measure is examined separately.

Next, we measured the potential underlying process. Participants rated how easy it was to *understand* what the speaker was saying (1 = very difficult to understand, 7 = very easy to understand) and completed two items of *competence* from prior work (i.e., Thompson and Ince 2013; “how competent is the speaker?” 1 = not at all competent, 7 = very competent; “how knowledgeable is the speaker?” 1 = not at all knowledgeable, 7 = very knowledgeable; *r* = .73). To test an alternative explanation, we also collected measures of movement typicality adapted from prior work (i.e., Kronrod, Grinstein, and Wathieu 2012; “this way of moving hands is typical,” “this way of moving hands is usual,” “I’ve often seen speakers moving their hands in this way” 1 = not at all, 7 = very; α = .90).

Finally, participants completed two additional measures for exploratory purposes (see details below), three attention checks, and demographic questions.

**Results**

*Persuasion*. As predicted, and consistent with Study 2, an ANOVA revealed a significant effect of movement type on persuasion (F(3, 681) = 3.67, *p* = .012). Specifically, pairwise comparisons (Figure 5) found that using illustrators (M = 4.22) boosted persuasion more than highlighters (M = 3.82, *b* = .42, SE = .14, *t* = 2.94, *p* = .003), unrelated movements (M = 3.92, *b* = .33, SE = .15, *t* = 2.24, *p* = .025) or no movement at all (M = 3.87, *b* = .38, SE = .14, *t* = 2.74, *p* = .006). Examining each outcome variable separately (i.e., speaker liking, product liking, and purchase intention) reveals the same results (see Web Appendix B).

*Understandability.* As predicted, movement type also influenced understandability (F(3, 681) = 9.78, *p* < .001). Pairwise comparisons found that using illustrators (M = 6.37) made the content easier to understand than highlighters (M = 6.02, *b* = .34, SE = .13, *t* = 2.65, *p* = .008), unrelated movements (M = 5.75, *b* = .61, SE = .13, *t* = 4.72, *p* < .001) or no movement at all (M = 5.79, *b* = .57, SE = .12, *t* = 4.61, *p* < .001).

*Perceived Competence.* As expected, similar effects were found for competence (F(3, 681) = 6.95, *p* < .001). Pairwise comparisons found that using illustrators (M = 5.34) made the speaker seem more competent than when they used highlighters (M = 4.94, *b* = .37, SE = .13, *t* = 2.81, *p* = .005), unrelated movements (M = 4.84, *b* = .46, SE = .13, *t* = 3.50, *p* < .001) or no movement at all (M = 4.79, *b* = .54, SE = .13, *t* = 4.27, *p* < .001).

**Figure 5**: Persuasion in Movement Type Conditions

*Note*: Error bars represent standard errors of the mean.

*Mediation*. Serial mediation with 5,000 bootstrap resamples (Preacher and Hayes 2004) found that understandability and perceived competence serially mediated illustrator’s effects (e.g., compared to highlighters: indirect effect = .13, 95% Confidence Interval (CI) = .03, .22). Illustrators made content easier to understand (*b* = .34, SE = .11, *t* = 2.27, *p* = .003), which made the speaker seem more competent (*b* = .53, SE = .05, *t* = 10.46, *p* < .001), which boosted persuasion (*b* = .70, SE = .05, *t* = 13.07, *p* < .001). Including these mediators led the conditional direct effect to be reduced to non-significance (95% CI = –.05, .40). Comparing illustrators to unrelated movements, or no movement, finds the same results (see Web Appendix B).

*Alternative Explanation.* Rather than being driven by the hypothesized process, one could wonder whether the effects are driven by how typical the movements seem. Maybe illustrators are simply more typical or familiar, and that encourages positive evaluations. Consistent with the field data (i.e., where illustrators were the least frequent movement), however, typicality did not vary by condition (F(3, 681) = 1.40, *p* = .243), casting doubt on this alternative.

**Discussion**

Study 3a provides direct causal support for our theorizing. Consistent with Study 2, illustrators boosted communication’s persuasive impact. They made audiences evaluate the speaker and product discussed more favorably and increased their interest in purchasing it.

Further, the results highlight a mechanism underlying the effects. Using illustrators made content easier to understand, which made communicators seem more competent, which boosted persuasion (Study 3b tests other potential accounts).

In addition, the results cast doubt on an alternative explanation based on movement typicality, and the fact that the effects hold in a different context (i.e., entrepreneur’s pitch) speaks to their generalizability.

*Exploratory Analyses*. One could wonder whether illustrators make spoken content more concrete or vivid. While this is essentially a manipulation check (i.e., consistent with the conceptualization of movement types, illustrators should provide a more concrete and tangible image of the spoken content compared to other movement types), to test this possibility, participants were asked how much the speaker created a concrete/vivid/tangible image of what they said (1 = not at all, 7 = very; α = .93).[[17]](#footnote-18)

One could also wonder whether audience expertise moderated the effect. To test this possibility, participants rated “how knowledgeable are you about that kind of product? (1 = not at all knowledgeable, 7 = very knowledgeable).

Results (see Web Appendix B for details) indicate that while illustrators made spoken content more concrete and vivid, audience expertise did not moderate the effects.

**Study 3b: Testing Robustness**

Study 3b (N = 892, Prolific; <https://aspredicted.org/WYG_HYN>) tests robustness in several ways. First, one might wonder whether the results are driven by the specific set of stimuli used in Study 3a. Consequently, we test whether they hold across multiple different pitches (stimulus-sampling; Wells and Windschitl 1999) and types of movement (i.e., the exact gesture used in each recording varied across the products). Second, while we suggested that illustrators boost persuasion because they make the content easier to understand, which makes the speaker seem more competent, multiple processes may contribute to the effects. Maybe illustrators make it seem like the speaker is more extraverted (or emotionally involved), genuine, or cares more about audience understanding. Alternatively, illustrators might increase attention, encourage observers imagine using the product, or be more engaging to watch. Consequently, in addition to understanding and competence, Study 3b measures these other possibilities, testing whether they mediate illustrators’ impact. Third, we explore whether the effects are robust to an alternate measure of competence (i.e., confidence with the subject matter; Fiske et al. 2002; Wang et al. 2021), another outcome variable (i.e., word-of-mouth), and further whether movement typicality can explain the effects. Given space constraints, we report the results briefly, but see Web Appendix B for complete details.

Consistent with prior studies, illustrators increased persuasion (MIllustrator = 4.92 vs. MHighlighter = 4.60, *b* = .32, SE = .13, *t* = 2.38, *p* = .017), unrelated movements (MIllustrator = 4.92 vs. MUnrelated = 4.60, *b* = .32, SE = .14, *t* = 2.32, *p* = .020) or no movement at all (MIllustrator = 4.92 vs. MNo movement = 4.56, *b* = .36, SE = .13, *t* = 2.73, *p* = .006). Further, understandability and perceived competence serially mediated illustrator’s effects compared to highlighters (albeit marginally, indirect effect: *p* = .065, 95% CI = –.003, .11), unrelated movements (albeit marginally, indirect effect: *p* = .051, 95% CI = –.0003, .11), and no movement (indirect effect: 95% CI = .001, .10).[[18]](#footnote-19) Results also reveal that illustrators have other benefits. Relative to unrelated movements or no movement, for example, illustrators increased attention, and relative to highlighters, they increased perceived caring (see Table WB1). That said, making speakers seem more competent, by making the content easier to understand, was the *only* process that explained illustrators’ positive impact relative to *all* three other movement types.

**Study 4a: Process by Moderation**

To further test the process underlying illustrators’ effects, we return to the field data. If one reason illustrators boost evaluations is because they make content easier to understand (which makes the speaker seem more competent), as we have suggested, then their effects should be mitigated when that content is easy to understand (i.e., simpler) to begin with. Given talking about having a sip of water is relatively easy to understand, for example, making a drinking gesture should have less of a positive impact on understandability, and thus illustrators should have less of an effect.

To test this possibility, Study 4a examines the moderating role of language complexity (i.e., “mediation-via-moderation”; Bullock, Green, and Ha 2010). If understandability is playing a role, illustrators’ positive effects should be mitigated when the content (i.e., language) is simpler. Following prior work (e.g., Skierkowski et al. 2019), complexity was measured using the SMOG index – a well-established metric that estimates the years of education required to comprehend a given text (Laughlin and Harry 1969). A sentence like “This drink is very cold,” for example, has a lower score than “This drink undergoes thermodynamic stabilization” (3.13 vs. 13.02), reflecting less complexity. SMOG scores (mean = 11.35, SD = 1.75, ranging from 6.68 to 35.95) were normalized (mean = 0, SD = 1) to facilitate interpretation.

As predicted, results reveal a significant illustrator × complexity interaction (*b* = .018, SE = .009, *t* = 2.09, *p* = .037).[[19]](#footnote-20) Consistent with the underlying role of understandability, floodlight analysis (Spiller et al. 2013) reveals that the positive effect of illustrators was mitigated for lower levels of language complexity (i.e., SMOG < –0.25, Figure 6).

**Figure 6**: Moderation by ComplexityA graph with blue lines

Description automatically generated

*Notes*: The dashed line indicates the level of complexity (Johnson-Neyman point: –0.25) at which illustrators starts being beneficial (right region).

**Discussion**

Study 4a provides further insight into why illustrators might be beneficial. Consistent with the underlying role of understandability, and findings from Studies 3a and 3b, Study 4 demonstrates that the illustrators’ positive impact was mitigated when what was being talking about was simpler to begin with.

*Study 4b*. An additional study further tests the underlying role of competence, testing the moderating role of language certainty (see Web Appendix C).

**General Discussion**

From salespeople and entrepreneurs to politicians and educators, marketplace actors often communicate with their hands. But while it’s clear that what communicators say affects their impact, less is known about whether the way they move their hands while speaking might also play a role.

The present research begins to address this gap. A multimethod investigation, combining field data with a preregistered experiment, demonstrates hand movement’s impact, identifies which gestures are more effective, and illustrates a mechanism underlying these effects.

First, the results shed light on the effect of hand movements. Automated video analysis of thousands of presentations (Study 1) finds that presentations are evaluated more positively when speakers move their hands more. Casting doubt on alternative explanations, the effects persist even accounting for aspects of the speaker, content, video, and presentation itself.

Second, the results demonstrate that certain movements are more impactful than others. A large multimodal model (Study 2) automatically identifies different types of movements (i.e., illustrators, highlighters, and unrelated movements) and finds that presentations are liked more when more illustrators are used. Experimentally manipulating movement type (Studies 3a and 3b) underscores illustrators’ causal impact and demonstrates that they boost persuasion (i.e., speaker evaluations and interest in buying the product discussed).

Third, using both mediation (Studies 3a and 3b) and moderation (Studies 4a and 4b) the results shed light on one reason why illustrators are beneficial. Illustrators make content easier to understand, which makes speakers seem more confident, which has a range of positive consequences (Studies 3a and 3b). Further, consistent with the notion that the effects are driven by understandability and competence, they are mitigated when the content is easier to understand (i.e., the language is less complex, Study 4a) and when other competence cues are present (using certainty-oriented language, Study 4b).

Finally, the fact that the results hold across hundreds of different products and ideas, gesture variants, outcomes (i.e., speaker liking, product liking, and purchase intent), and domains (i.e., expert talks and entrepreneur pitches) speaks to their generalizability.

**Contributions**

This work makes several contributions. First, it highlights how body language can shape consumer behavior. A burgeoning stream of research has begun to examine how the words and vocal cues communicators use influence their impact (e.g., Berger et al. 2020; Luangrath, Xu, and Wang 2023; Sela, Wheeler, and Sarial-Abi 2012; Wang et al. 2021; Wang et al. 2024). But while these studies have provided many valuable insights, less attention has been paid to the role of body language. Our findings demonstrate the impact of hand movement. Given how many interactions involve visual communication (e.g., face-to-face or through video), this area deserves further attention.

Second, we advance understanding of how different communication modes combine to shape impact. Communication often involves a blend of verbal, visual, and auditory elements, but most extant work has examined these modes separately, focusing on either linguistic (e.g., Berger et al. 2020) or visual elements (e.g., Hartmann et al. 2021; He, Li, and Wang 2023). The current research explores the integration of spoken language and hand gestures, advancing beyond words or visual displays to deeper reasoning about how they relate to one other. In doing so, it contributes to work (e.g., Ceylan, Diehl, and Proserpio 2024) examining how different modes interact. Given that much of today’s content (e.g., TV shows and social media videos) is inherently multimodal, understanding these interactions is fundamental.

Third, methodologically, we highlight the value of automated video analysis in marketing research. While more and more work has started using natural language processing (e.g., Berger et al. 2020; Humphreys and Wang 2018; Li, Shi, and Wang 2019), there’s been less attention to computer vision (e.g., Hartmann et al. 2021; Zhang et al. 2024). Further, the studies that have been conducted mainly focused on images rather than videos (c.f., Lu, Xao, and Ding 2016; Zhou et al. 2021). Given the prevalence of video across contexts (e.g., interactions with people, firms, and technology) and forms of consumption (e.g., TikTok videos and live-streamed events), this area deserves deeper exploration.

Fourth, the present research also provides guidelines for using LMMs in behavioral research. While large language models (LLMs) have potential in behavioral research (Arora, Chakraborty, and Nishimura 2024), their “black box” nature raises concerns about accuracy and reliability (Li et al. 2024). We offer a systematic pipeline for leveraging LMMs to measure and validate constructs. We developed a framework using Gemini LMM to classify hand movements in videos, linking gestures with spoken content. The process involved several steps, from defining coding guidelines and training human annotators to prompting Gemini with examples and refining the model’s accuracy through iterative feedback. This systematic approach addresses the limitations of text- or visual-only models and provides a practical method for leveraging LMMs in behavioral research, aiming to stimulate further work in the field. In addition, note that while papers have analyzed text using language AI (i.e., large language models; Chakraborty, Kim, and Sudhir 2022) or images using text-to-image models (e.g., Zhou and Lee 2024), to the best of our knowledge, this is the first marketing paper examining the relationship between text and visuals using language-vision AI (i.e., large multimodal models).

**Practical Implications**

The findings have important practical implications. The results suggest that a relatively simple adjustment – moving one’s hands more, and in particular using more illustrators – can have beneficial effects. The field data, for example, suggest that doubling hand movement is associated with a 5.16% increase in liking. Compared to more resource-intensive strategies (e.g., programmatic advertising), adjusting hand movement seems easier and potentially more cost-effective to implement. Despite these benefits, only 6.5% of the hand movements used during a presentation were illustrators, suggesting underutilized communication strategy.

These insights can be valuable for a range of marketplace actors. Brands and content creators, for example, increasingly rely on video platforms (e.g., TikTok or Twitch) to sell products (e.g., livestreaming commerce). Increasing hand movement can be a simple way to boost audience engagement and willingness to purchase. Politicians, educators, and corporate leaders can also benefit. In conferences, classrooms, or board meetings, where effective communication is critical, hand movements—specifically illustrators—can make speakers seem more competent and increase their impact.

Crucially, results indicate that this skill can be easily taught and adopted. As shown in Studies 3a and 3b, even minimally trained speakers were able to change their hand movements to achieve the desired effect. This highlights that the strategic use of hand movements is **teachable**, **scalable**, and applicable across a range of communicators.

The findings also have implications for human-computer interaction. As AI-driven avatars and virtual assistants become more integrated into everyday life (e.g., “virtual influencers”), their ability to replicate human-like communication becomes a key factor in user acceptance and engagement (Puntoni et al. 2021). These results suggest that using illustrators could lead such AI-driven avatars to be perceived more positively.

**Limitations and Future Research**

As with any emerging area, these results raise questions for future research. First, individual differences, cultural norms, and situational factors might moderate the effects. Cultures vary in gesture use (Kita 2009). In Western cultures, for example, hand gestures are used liberally to emphasize points and convey enthusiasm, while in Eastern cultures they might be reserved for formal or specific contexts (reflecting a preference for more restrained and controlled nonverbal communication; Matsumoto and Hwang 2013). Research could investigate how the contextual use of gestures influences their effectiveness.

Second, work might examine how speakers select what types of gestures to use. If someone doesn’t know much about a topic, for example, they might have trouble using illustrators. Gesture type might also depend on what speakers are talking about. For complex topics, speakers might be more likely to use illustrators (Dargue, Sweller, and Jones 2019).

Third, not all illustrators may equally signal competence. While illustrators can enhance understandability, those that visually represent abstract or complex ideas (i.e., metaphorical illustrators) may do more to signal competence. Translating abstract concepts into tangible visual forms likely requires greater fluency and communicative skill, making such gestures more diagnostic of expertise in the eyes of observers. By contrast, simpler iconic gestures (e.g., showing the size of an object while saying “I caught a big fish”) may primarily aid clarity without implying deep subject mastery. Further, while some work suggests that being able to deal with complex topics can signal understanding (Thompson and Ince 2013), our findings highlight that being able to use metaphorical gestures indicates that the speaker understands the topic deeply enough to explain it clearly. Future research could examine whether, and under what conditions, metaphorical and iconic illustrators differentially influence perceptions of competence.

Third, we focused on hand movements, but other forms of nonverbal communication also deserve attention. Some work has begun to examine some aspects of kinesics (e.g., facial expressions; Zhang et al. 2024) and haptics (i.e., interpersonal touch; Luangrath, Peck, and Gustafsson 2020), yet little is known about proxemics (i.e., interpersonal physical distance). Psychological research suggests that increasing spatial distance can be used by individuals in higher status positions to assert authority (Burgoon, Birk, and Pfau 1990), while reducing distance signals extraversion (Kroencke et al. 2023) and a desire for intimacy (Fay and Maner 2012). Future studies could investigate how aspects of proxemics influence others’ behavior.

Finally, while the effects in Studies 3a and 3b are statistically robust, the manipulations (i.e., brief clips with a single gesture) were intentionally minimal. This offers a **conservative test** of our theorizing and illustrates what is for gestures to shape impact. The modest size of some effects, particularly in Study 3b, may also reflect limited statistical power, especially given the subtlety of the manipulation and potential constraints in video visibility. Future work could also explore whether more exaggerated movement backfires. While Study 1 found no evidence of a quadratic effect, TED Talks may not capture extreme gestures, leaving open the possibility of nonlinear effects in other settings.

**Conclusion**

In conclusion, this research demonstrates how communicators move their hands while speaking can have important impacts on evaluations and persuasion. In doing so, it sheds light on the effects of body language in the marketplace, and on consumer behavior more broadly.

**References**

Abrams, Richard A. and Shawn E. Christ (2003), “Motion onset captures attention,” *Psychological Science*, 14 (5), 427-432.

Alibali, Martha Wagner and Alyssa A. DiRusso (1999), “The function of gesture in learning to count: more than keeping track,” *Cognitive Development,* 14, 37–56.

Arora, Neeraj, Ishita Chakraborty, and Yohei Nishimura (2024), “AI-Human Hybrids for Marketing Research: Leveraging LLMs as Collaborators,” *Journal of Marketing*, forthcoming.

Asalıoğlu, Esma Nur and Tilbe Göksun (2023), “The role of hand gestures in emotion communication: Do type and size of gestures matter?,” *Psychological Research*, 87 (6), 1880-1898.

Austin, Elizabeth E., Naomi Sweller, and Penny Van Bergen (2018), “Pointing the way forward: Gesture and adults’ recall of route direction information,” *Journal of Experimental Psychology: Applied*, 24 (4), 490.

Beattie, Geoffrey and Heather Shovelton (1999), “Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation,” *Semiotica*, 123–1/2, 001–030.

Beichert, Maximilian, Andreas Bayerl, Jacob Goldenberg, and Andreas Lanz (2024), “Revenue generation through influencer marketing,” *Journal of Marketing*, 88 (4), 40-63.

Beilock, Sian L. and Susan Goldin-Meadow (2010), “Gesture changes thought by grounding it in action,” *Psychological Science,* 21, 1605–1611.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy Moe, Oded Netzer, and David A. Schweidel (2020), “Uniting the tribes: Using text for marketing insight,” *Journal of Marketing*, 84 (1), 1-25.

Briñol, Pablo, Richard E. Petty, David Santos, and Joana Mello (2017), “Meaning moderates the persuasive effect of physical actions: Buying, selling, touching, carrying, and cleaning thoughts as if they were commercial products,” *Journal of the Association for Consumer Research*, 2 (4), 460-471.

Bullock, John G., Donald P. Green, and Shang E. Ha (2010), “Yes, but what’s the mechanism? (don’t expect an easy answer),” *Journal of Personality and Social Psychology*, 98 (4), 550.

Burgoon, Judee K., Thomas Birk, and Michael Pfau (1990), “Nonverbal behaviors, persuasion, and credibility,” *Human Communication Research*, 17 (1), 140-169.

Burgoon, Judee K., David B. Buller, and William Gill Woodall (1996), *Nonverbal communication: The unspoken dialogue*, New York: McGraw-Hill.

Burgoon, Judee K., Ryan Schuetzler, and David W. Wilson (2015), “Kinesic patterning in deceptive and truthful interactions,” *Journal of Nonverbal Behavior*, 39, 1-24.

Burgoon, Judee K., Valerie Manusov, and Laura K. Guerrero (2021), *Nonverbal communication*, Routledge.

Carlson, Emily, Birgitta Burger, Justin London, Marc R. Thompson, and Petri Toiviainen (2016), “Conscientiousness and extraversion relate to responsiveness to tempo in dance,” *Human movement science*, 49, 315-325.

Carney, Russell N. and Joel R. Levin (2002), “Pictorial illustrations still improve student’' learning from text,” *Educational Psychology Review*, 14, 5-26.

Cascio Rizzo, Giovanni Luca, Jonah Berger, Matteo De Angelis, and Rumen Pozharliev (2023), “How sensory language shapes influencer’s impact,” *Journal of Consumer Research*, 50 (4), 810-825.

Cascio Rizzo, Giovanni Luca, Francisco Villarroel Ordenes, Rumen Pozharliev, Matteo De Angelis, and Michele Costabile (2024), “How high-arousal language shapes micro-versus macro-influencers’ impact,” *Journal of Marketing*, 80 (4), 107-128.

Caso, Letizia, Fridanna Maricchiolo, Marino Bonaiuto, Aldert Vrij, and Samantha Mann (2006), “The impact of deception and suspicion on different hand movements,” *Journal of Nonverbal Behavior*, 30, 1-19.

Cesario, Joseph and E. Tory Higgins (2008), “Making message recipients “feel right” how nonverbal cues can increase persuasion,” *Psychological Science*, 19 (5), 415-420.

Ceylan, Gizem, Kristin Diehl, and Davide Proserpio (2024), “Words meet photos: When and why photos increase review helpfulness,” *Journal of Marketing Research*, 61 (1), 5-26.

Chaiken, Shelly and Durairaj Maheswaran (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66 (3), 460.

Chakraborty, Ishita, Minkyung Kim, and K. Sudhir (2022), “Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes,” *Journal of Marketing Research*, 59 (3), 600-622.

Chi, Michelene T., Paul J. Feltovich, and Robert Glaser (1981), “Categorization and representation of physics problems by experts and novices,” *Cognitive Science*, 5 (2), 121-152.

Chung, Jaeyeon, Yu Ding, and Ajay Kalra (2023), “I really know you: how influencers can increase audience engagement by referencing their close social ties,” *Journal of Consumer Research*, 50 (4), 683-703.

Clarke, Jean S., Joep P. Cornelissen, and Mark P. Healey (2019), “Actions speak louder than words: How figurative language and gesturing in entrepreneurial pitches influences investment judgments,” *Academy of Management Journal*, 62 (2), 335-360.

Cohen, Jacob, Patricia Cohen, Stephen G. West, and Leona S. Aiken (2013), *Applied multiple regression/correlation analysis for the behavioral sciences*, Routledge.

Cook, Michelle Patrick (2006), “Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles,” *Science  
Education*, 90 (6), 1073-1091.

Cook, Susan Wagner, Ryan G. Duffy, and Kimberly M. Fenn (2013), “Consolidation and transfer of learning after observing hand gesture,” *Child Development*, 84 (6), 1863-1871.

Cook, Susan Wagner, Howard S. Friedman, Katherine A. Duggan, Jian Cui, and Voicu Popescu (2017), “Hand gesture and mathematics learning: lessons from an Avatar,” *Cognitive Science*, 41 (2), 518-535.

Dael, Nele, Marcello Mortillaro, and Kraus R. Scherer (2012), “Emotion expression in body action and posture,” *Emotion*, 12 (5), 1085.

Dargue, Nicole and Naomi Sweller (2018), “Not all gestures are created equal: The effects of typical and atypical iconic gestures on narrative comprehension,” *Journal of Nonverbal Behavior*, 42, 327-345.

Dargue, Nicole, Naomi Sweller, and Michael P. Jones (2019), “When our hands help us understand: A meta-analysis into the effects of gesture on comprehension,” *Psychological Bulletin*, 145 (8), 765.

Dubois, David, Derek D. Rucker, and Adam D. Galinsky (2016), “Dynamics of communicator and audience power: The persuasiveness of competence versus warmth,” *Journal of Consumer Research*, 43 (1), 68-85.

Ekman, Paul and Wallace V. Friesen (1972), “Hand movements,” *Journal of Communication*, 22 (4), 353-374.

Ekman, Paul and Wallace V. Friesen (1974), “Detecting deception from the body or face,” *Journal of Personality and Social Psychology*, 29 (3), 288.

Escalas, Jennifer Edison (2004), “Narrative processing: Building consumer connections to brands,” *Journal of Consumer Psychology*, 14 (1-2), 168-180.

Fay, Adam J. and Jon K. Maner (2012), “Warmth, spatial proximity, and social attachment: The embodied perception of a social metaphor,” *Journal of Experimental Social Psychology*, 48 (6), 1369-1372.

Fiske, Susan T., Amy JC Cuddy, Peter Glick, and Jun Xu (2002), “A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition,” *Journal of Personality and Social Psychology*, 82 (6), 878–902.

Fiske, Susan T., Amy JC Cuddy, and Peter Glick (2007), “Universal dimensions of social cognition: Warmth and competence,” *Trends in Cognitive Sciences*, 11 (2), 77-83.

Fiske, Susan T. (2017), “Prejudices in cultural contexts: Shared stereotypes (gender, age) versus variable stereotypes (race, ethnicity, religion),” *Perspectives on psychological science*, *12*(5), 791-799.

Franconeri, Steven L. and Daniel J. Simons (2003), “Moving and looming stimuli capture attention,” *Perception & Psychophysics*, 65 (7), 999-1010.

Garcia-Retamero, Rocio and Edward T. Cokely (2017), “Designing visual aids that promote risk literacy: A systematic review of health research and evidence-based design heuristics,” *Human Factors*, 59 (4), 582-627.

Gluhareva, Daria and Pilar Prieto (2017), “Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations,” *Language Teaching Research*, 21 (5), 609-631.

Goldin-Meadow, Susan (1999), “The role of gesture in communication and thinking,” *Trends in Cognitive Sciences*, 3 (11), 419-429.

Goldin-Meadow, Susan and Sian L. Beilock (2010), “Action’s influence on thought: The case of gesture,” *Perspectives on Psychological Science,* 5, 664–674.

Goldin-Meadow, Susan and Martha Wagner Alibali (2013), “Gesture’s role in speaking, learning, and creating language,” *Annual Review of Psychology*, 64 (1), 257-283.

Google DeepMind (2024), “Gemini Models,” https://deepmind.google/technologies/gemini/pro/.

Grewal, Rajdeep, Sachin Gupta, and Rebecca Hamilton (2021), “Marketing insights from multimedia data: text, image, audio, and video,” *Journal of Marketing Research*, 58 (6), 1025-1033.

Hartmann, Jochen, Mark Heitmann, Christina Schamp, and Oded Netzer (2021), “The power of brand selfies,” *Journal of Marketing Research*, 58 (6), 1159-1177.

He, Jiaxiu, Bingqing Li, and Xi (Shane) Wang (2023), “Image features and demand in the sharing economy: A study of Airbnb,” *International Journal of Research in Marketing*, 40 (4), 760-780.

Hildebrand, Christin, Fotis Efthymiou, Francesc Busquet, William H. Hampton, Donna L. Hoffman, and Thomas P. Novak (2020), “Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications,” *Journal of Business Research*, 121, 364-374. Hoogervorst, Chiara (2023), “Brand Videos: The Key to Great Video Marketing in 2025,” Wistia (accessed May 26th, 2025), <https://wistia.com/learn/marketing/creative-brand-videos>.

Hostetter, Autumn B. (2011), “When do gestures communicate? A meta-analysis,” *Psychological Bulletin*, *137*(2), 297.

Hostetter, Autumn B. and Andrea L. Potthoff (2012), “Effects of personality and social situation on representational gesture production,” *Gesture*, 12 (1), 62-83.

Huang, Karen, Michael Yeomans, Alison Wood Brooks, Julia Minson, and Francesca Gino (2017), “It doesn’t hurt to ask: Question-asking increases liking,” *Journal of Personality and Social Psychology*, 113 (3), 430.

Humphreys, Ashlee (2010), “Semiotic structure and the legitimation of consumption practices: The case of casino gambling,” *Journal of Consumer Research*, 37 (3), 490-510.

Humphreys, Ashlee and Rebecca Jen-Hui Wang (2018), “Automated text analysis for consumer research,” *Journal of Consumer Research*, 44 (6), 1274-1306.

Humphreys, Ashlee, Mathew S. Isaac, and Rebecca Jen-Hui Wang (2021), “Construal matching in online search: Applying text analysis to illuminate the consumer decision journey,” *Journal of Marketing Research*, 58 (6), 1101-1119.

Kang, Seokmin and Barbara Tversky (2016), “From hands to minds: Gestures promote understanding,” *Cognitive Research: Principles and Implications*, 1, 1-15.

Kelly, Spencer, Dale J. Barr, R. Breckinridge Church, and Katheryn Lynch (1999), “Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory,” *Journal of Memory and Language*, 40 (4), 577-592.

Kendon, Adam (2004), “Gesture: Visible action as utterance,” Cambridge University Press.

Kita, Sotaro (2009), “Cross-cultural variation of speech-accompanying gesture: A review,” *Language and Cognitive Processes*, 24 (2), 145-167.

Kroencke, Lara, Sarah Humberg, Simon M. Breil, Katharina Geukes, Giulia Zoppolat, Rhonda Nicole Balzarini, Maria Alonso-Ferres, Richard B. Slatcher, and Mitja Back (2023),” Extraversion, social interactions, and well-being during the COVID-19 pandemic: Did extraverts really suffer more than introverts?,” *Journal of Personality and Social Psychology*, 125 (3), 649.

Kronrod, Ann, Amir Grinstein, and Luc Wathieu (2012), “Enjoy! Hedonic consumption and compliance with assertive messages,” *Journal of Consumer Research*, 39 (1), 51-61.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P. K. Kannan (2016), “From social to sale: The effects of firm-generated content in social media on customer behavior,” *Journal of Marketing*, 80 (1), 7-25.

Lafreniere, Katherine C., Sarah G. Moore, and Robert J. Fisher (2022), “The power of profanity: The meaning and impact of swear words in word of mouth,” *Journal of Marketing Research*, 59 (5), 908-925.

Li, Xi, Mengze Shi, and Xin (Shane) Wang (2019), “Video mining: Measuring visual information using automatic methods,” *International Journal of Research in Marketing*, 36 (2), 216-231.

Li, Yiyi and Ying Xie (2020), “Is a picture worth a thousand words? An empirical study of image content and social media engagement,” *Journal of Marketing Research*, 57 (1), 1-19.

Li, Peiyao, Noah Castelo, Zsolt Katona, and Miklos Sarvary (2024), “Frontiers: Determining the validity of large language models for automated perceptual analysis,” *Marketing Science*, 43 (2), 254-266.

Liadeli, Georgia, Francesca Sotgiu, and Peter W. Verlegh (2023), “A meta-analysis of the effects of brands’ owned social media on social media engagement and sales,” *Journal of Marketing*, 87 (3), 406-427.

Lowe, Michael L. Kelly L. and Haws (2017), “Sounds big: The effects of acoustic pitch on product perceptions,” *Journal of Marketing Research*, 54 (2), 331-346.

Lu, Shasha, Li Xiao, and Min Ding (2016), “A Video-Based Automated Recommender (VAR) System for Garments,” *Marketing Science*, 35 (3), 484–510.

Luangrath, Andrea Webb, Joann Peck, and Victor A. Barger (2017), “Textual paralanguage and its implications for marketing communications,” *Journal of Consumer Psychology*, 27 (1), 98-107.

Luangrath, Andrea Webb, Joan Peck, and Anders Gustafsson (2020), “Should I touch the customer? Rethinking interpersonal touch effects from the perspective of the touch initiator,” *Journal of Consumer Research*, 47 (4), 588-607.

Luangrath, Andrea Webb, Yixiang Xu, and Tong Wang (2023), “Paralanguage classifier (PARA): An algorithm for automatic coding of paralinguistic nonverbal parts of speech in text,” *Journal of Marketing Research*, 60 (2), 388-408.

Lurie, Nicholas H. and Charlotte H. Mason (2007), “Visual representation: Implications for decision making,” *Journal of Marketing*, 71 (1), 160-177.

Matsumoto, David and Hyisung C. Hwang (2013), “Cultural similarities and differences in emblematic gestures,” *Journal of Nonverbal Behavior*, 37, 1-27.

McGregor, Karla, Katharina J. Rohlfing, Allison Bean, and Ellen Marschner (2009), “Gesture as support for word learning: The case of *under*,” *Journal of Child Language,* 36, 807–828.

McNeill, David (1992), “Hand and mind,” *Advances in Visual Semiotics*, 351.

Mc Laughlin, G. H. (1969), “SMOG grading-a new readability formula,” *Journal of Reading*, 12 (8), 639-646.

Mehrabian, Albert (1971), *Silent Messages*, Wadsworth.

Michail, Jon (2020), “Strong Nonverbal Skills Matter Now More Than Ever In This New Normal,” https://www.forbes.com/councils/forbescoachescouncil/2020/08/24/strong-nonverbal-skills-matter-now-more-than-ever-in-this-new-normal/.

Moore, Sarah G. (2012), “Some things are better left unsaid: How word of mouth influences the storyteller,” *Journal of Consumer Research*, 38 (6), 1140-1154.

Moore, Sarah G. (2015), “Attitude predictability and helpfulness in online reviews: The role of explained actions and reactions,” *Journal of Consumer Research*, 42 (1), 30-44.

Mortillaro, Marcello and Daniel Dukes (2018), “Jumping for joy: the importance of the body and of dynamics in the expression and recognition of positive emotions,” *Frontiers in Psychology*, 9, 763.

Mrkva, Kellen, Jacob Westfall, and Leaf Van Boven (2019), “Attention drives emotion: Voluntary visual attention increases perceived emotional intensity,” *Psychological Science*, 30 (6), 942-954.

Packard, Grant, Sarah G. Moore, and Brent McFerran (2018), “(’'m) happy to help (you): The impact of personal pronoun use in customer–firm interactions,” *Journal of Marketing Research*, 55 (4), 541-555.

Packard, Grant and Jonah Berger (2021), “How concrete language shapes customer satisfaction,” *Journal of Consumer Research*, 47 (5), 787-806.

Packard, Grant, Jonah Berger, and Reihane Boghrati (2023), “How verb tense shapes persuasion,” *Journal of Consumer Research*, 50 (3), 645-660.

Packard, Grant, and Jonah Berger (2024), “The Emergence and Evolution of Consumer Language Research,” *Journal of Consumer Research*, 51 (1), 42-51.

Paetzold, Gustavo and Lucia Specia (2016), “Inferring psycholinguistic properties of words,” In *Proceedings of the 2016 conference of the northAmericann chapter of the association for computational linguistics: Human language technologies* (pp. 435-440).

Paivio, Allan (1969), “Mental imagery in associative learning and memory,” *Psychological Review*, 76 (3), 241.

Pancer, Ethan, Vincent Chandler, Maxwell Poole, and Theodore J. Noseworthy (2019), “How readability shapes social media engagement,” *Journal of Consumer Psychology*, 29 (2), 262-270.

Parzuchowski, Michal and Bodgan Wojciszke (2014), “Hand over heart primes moral judgments and behavior,” *Journal of Nonverbal Behavior*, 38, 145-165.

Patrick, Vanessa M. and Henrik Hagtvedt (2012a), “I don’t” versus “I can’t”: When empowered refusal motivates goal-directed behavior,” *Journal of Consumer Research*, 39 (2), 371-381.

Patrick, Vanessa M. and Henrik Hagtvedt (2012b), “How to say “no”: Conviction and identity attributions in persuasive refusal,” *International Journal of Research in Marketing*, 29 (4), 390-394.

Pi, Zhongling, Kianzhong Hong, and Jiumin Yang (2017), “Effects of the instructo’'s pointing gestures on learning performance in video lectures,” *British Journal of Educational Technology*, 48 (4), 1020-1029.

Ping, Raedy M. and Susan Goldin-Meadow (2008), “Hands in the air: using ungrounded iconic gestures to teach children conservation of quantity,” *Developmental Psychology*, 44 (5), 1277.

Preacher, Kristopher J. and Andrew F. Hayes (2004), “SPSS and SAS procedures for estimating indirect effects in simple mediation models,” *Behavior Research Methods, Instruments, & Computers*, 36, 717-731.

Puntoni, Stefano, Rebecca Walker Reczek, Markus Giesler, and Simona Botti (2021), “Consumers and artificial intelligence: An experiential perspective,” *Journal of Marketing*, 85 (1), 131-151.

Remiro, Miguel Ángel, Manuel Gil-Martín, and Rubén San-Segundo (2023), “Improving Hand Pose Recognition Using Localization and Zoom Normalizations over MediaPipe Landmarks,” *Engineering Proceedings*, *58* (1), 69.

Rocklage, Matthew D., Derek D. Rucker, and Loran F. Nordgren (2018), “The Evaluative Lexicon 2.0: The measurement of emotionality, extremity, and valence in language,” *Behavior Research Methods*, 50, 1327-1344.

Sela, Aner, S. Christian Wheeler, and Gülen Sarial-Abi (2012), “We are not the same as you and I: Causal effects of minor language variations on consumer’' attitudes toward brands,” *Journal of Consumer Research*, 39 (3), 644-661.

Skierkowski, Dorothy D., Paul Florin, Lisa L. Harlow, Jason Machan, and Yinjiao Ye (2019), “A readability analysis of online mental health resources,” *American Psychologist*, 74 (4), 474.

Spiller, Stephen A., Gavan J. Fitzsimons, John G. Lynch Jr, and Gary H. McClelland (2013), “Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression,” *Journal of Marketing Research*, 50 (2), 277-288.

Thompson, Debora V. and Elise Chandon Ince (2013), “When disfluency signals competence: The effect of processing difficulty on perceptions of service agents,” *Journal of Marketing Research*, 50 (2), 228-240.

Vakunov, Andrey, Chuo-Ling Chang, Fan Zhang, George Sung, Matthias Grundmann, and Vaentin Bazarevsky (2020), “Mediapipe hands: On-device real-time hand tracking,” In *Workshop on Computer Vision for AR/VR* (Vol. 2, No. 4, p. 5).

Valsesia, Francesca, Davide Proserpio, and Joseph C. Nunes (2020), “The positive effect of not following others on social media,” *Journal of Marketing Research*, 57 (6), 1152-1168.

Van Zant, Alexander B. and Jonah Berger (2020), “How the voice persuades,” *Journal of Personality and Social Psychology*, 118 (4), 661.

Vrij, Aldert, Lucy Akehurst, and Paul Morris (1997), “Individual differences in hand movements during deception,” *Journal of Nonverbal Behavior*, 21, 87-102.

Wang, Xin (Shane), Shijie Lu, X. I. Li, Mansur Khamitov, and Neil Bendle (2021), “Audio mining: The role of vocal tone in persuasion,” *Journal of Consumer Research*, 48 (2), 189-211.

Wang, Xin (Shane), Jiaxiu He, David J. Curry, and Jun Hyun (Joseph) Ryoo (2022), “Attribute embedding: Learning hierarchical representations of product attributes from consumer reviews,” *Journal of Marketing*, 86 (6), 155-175.

Wang, Xin (Shane), Neil Bendle, and Yinjie Pan (2024), “Beyond text: Marketing strategy in a world turned upside down,” *Journal of the Academy of Marketing Science*, 52, 939-954.

Wang, Xin (Shane), Jun Hyun (Joseph) Ryoo, Margaret C. Campbell, and J. Jeffrey Inman (2024), “Unraveling Impact: Exploring Effects of Novelty in Top Consumer Research Journals,” *Journal of Consumer Research*, 51 (1), 169-179.

Wang, Ze, Huifang Mao, Yexin Jessica Li, and Fan Liu (2017), “Smile big or not? Effects of smile intensity on perceptions of warmth and competence,” *Journal of Consumer Research*, 43 (5), 787-805.

Wells, Gary L. and Paul D. Windschitl (1999), “Stimulus sampling and social psychological experimentation,” *Personality and Social Psychology Bulletin*, 25 (9), 1115-1125.

Winkielman, Piotr, David E. Huber, Liam Kavanagh, and Norbert Schwarz (2012), “Fluency of consistency: When thoughts fit nicely and flow smoothly,” *Cognitive consistency: A fundamental principle in social cognition*, 89-111.

Zhang, Shunyuan, Elizabeth Friedman, Kannan Srinivasan, Ravi Dhar, and Xupin Zhang (2024), “Serving with a smile on Airbnb: Analyzing the economic returns and behavioral underpinnings of the host’s smile,” *Journal of Consumer Research*, ucae049.

Zhou, Eric and Dokyun Lee (2024), “Generative artificial intelligence, human creativity, and art,” *PNAS Nexus*, 3 (3), pgae052.

Zhou, Mi, George H. Chen, Pedro Ferreira, and Michael D. Smith (2021), “Consumer behavior in the online classroom: Using video analytics and machine learning to understand the consumption of video courseware,” *Journal of Marketing Research*, 58 (6), 1079-1100.

**WEB APPENDIX**

**Web Appendix A**

**Study 1**

**Movement Normalization**

To ensure accurate calculation of hand movements in video frames, we normalized hand landmark coordinates using a computed scale factor, that corrects for variations in camera distance and zoom levels. This scale factor was derived from the Euclidean distance between the wrist (landmark[0]) and the index fingertip (landmark[8]), providing a consistent measure of scale. For each frame, we multiplied the landmark coordinates by this scale factor, transforming them into a normalized space where movements can be consistently measured. Specifically, we calculated the Euclidean distance between corresponding landmarks in consecutive frames using these normalized coordinates. Because MediaPipe tracks only hand keypoints, and our normalization anchors on the wrist, the measure isolates hand movement rather than broader body motion (e.g., walking ot torso shift). To the best of our knowledge, this approach has not been previously applied in behavioral research and represents a novel use of MediaPipe for isolating and quantifying hand movement magnitude over time.

**Control Variables**

*Aspects of the Speaker.* Rather than being driven by hand movement, one could argue that the results are driven by aspects of the speaker. Certain professions or demographics might appeal to a wider audience, for example, so we control for speaker occupation (e.g., activist, inventor, or scientist; see Table WA2), gender (using ChatGPT 4o), (estimated) age and skin tone both computed using Face++ [(https://console.faceplusplus.com/documents/567894](file:///C:\Users\jberger\AppData\Local\Microsoft\Windows\INetCache\Content.Outlook\9KBIE1IT\(https:\console.faceplusplus.com\documents\567894)8).

*Aspects of Content*. What speakers say, how they say it, and their facial expressions also likely play a role, so we control for these aspects. First, one could wonder whether certain topics are received more positively, and this, rather than movement, is driving the effect. To address this, we control for the topics discussed (e.g., urban planning or AI) using latent Dirichlet allocation (LDA; Blei 2012). LDA captures the mixture of words that co-occur within and across talks to identify the main topics discussed (e.g., a given talk may be 30% about urban planning and 40% about AI). We identified the lowest number of topics that maximize predictive power (i.e., 20 topics as identified by coherence metric), and control for the proportion of each topic in each talk. Second, maybe people use their hands more when expressing emotion, and emotionality, rather than hand movement per se, is driving things, so we control for language emotionality using the Evaluative Lexicon (Rocklage, Rucker, and Nordgren 2018). Third, hand movements may make thoughts more concrete (or “visible”; Beattie 2004), and maybe people who move their hands more also use more concrete language, so we control for linguistic concreteness using Paetzold and Specia’s (2016) concreteness ratings. Fourth, maybe communicators use their hands to help tell a story, and maybe narrativity is driving the results. To test this possibility, we control for narrativity using Linguistic Inquiry and Word Count’s (LIWC; Boyd et al. 2022) “narrative arc” dimension. Fifth, maybe communicators use their hands more when talking about things that happened previously, and past experiences are evaluated differently, so we control for it using the LIWC “focus past” dimension. Sixth, questions can enhance liking (Huang et al. 2017), so we control for the proportion of questions using LIWC. Seventh, maybe people who use their hands more also employ more positive language, so we control for language positivity using LIWC “Tone” dimension. Eighth, familiar words are easier to process, which can increase liking (Pancer et al. 2019), so we control for this using Paetzold and Specia’s (2016) familiarity ratings. Ninth, swear words can convey meaning, which may increase persuasion (Lafreniere, Moore, and Fisher 2022), so we control for them using LIWC. Tenth, function words (e.g., pronouns or adverbs) provide insight into the psychological state of the speaker, which can increase persuasion (Sela, Wheeler, and Sarial-Abi 2012), so we use LIWC to control for them. Eleventh, we account for other major language features by controlling for the LIWC main dimensions (i.e., analytic, clout, authentic) and psychological process dictionaries (i.e., cognition, drives, social, time orientation, conversation).

We also collected vocal features. For each video, we removed background music and noise using Spleeter (<https://github.com/deezer/spleeter>) and then measured six vocal cues. Faster speakers might seem more competent and confident, which can increase persuasion (Cesario and Higgins 2008), so we control for speaking rate (average words per minute) using subtitles. Low-pitched voices are perceived as more confident, which can increase persuasion (e.g., Guyer, Fabrigar, and Vaughan-Jonhston 2021), so we control for both vocal pitch and pitch variation (pitch SD) using Praat’s Parsemouth. Individuals who speak louder are also seen as more competent, which can increase persuasion (Van Zant and Berger 2020), so we control for both voice volume and volume variation (volume SD). Higher values of Harmonics-to-Noise Ratio, which people perceive in voice quality, have been linked to increased confidence (Hildebrand et al. 2020), so we control for it using Praat’s Parsemouth. Similarly, vocal shimmer has been associated with vocal instability, stress, and confidence, so we control for it using Praat.

Finally, facial emotions can influence consumer evaluations (Li and Xie 2020), so following prior work (Hu and Ma 2024) we used Face++ to control for speakers’ facial emotional states (i.e., anger, disgust, fear, happiness, neutral, sadness, and surprise).

*Aspects of the Video.* Beyond who is presenting, or aspects of the content, aspects of the video may also affect evaluations. First, longer videos may be more likely to include movement, or just include more content, so we control for video duration (in minutes). Second, movement’s impact may depend on how hands are framed (e.g., if the hands are framed more prominently, movements may be more noticeable and influence evaluations), so following prior work (Zhou et al. 2021), we used an OpenCV algorithm to control for motion area (i.e., the percentage of the total frame occupied by the moving region), magnitude (i.e., mean displacement across all moving pixels in a frame), and direction (i.e., angle of the displacement vector).[[20]](#footnote-21) Third, changes in framing can make videos more enjoyable, so we control for average scene length using PySceneDetect (e.g., Zhou et al. 2021). Finally, visual aesthetic features like warm hue proportion (i.e., portion of pixels in warm colors in a frame), color saturation (i.e., average saturation values across all pixels in a frame), brightness (i.e., average intensity values across all pixels in a frame), contrast of brightness (i.e., standard deviation of pixel intensity across the whole frame), or clarity (i.e., portion of pixels with brightness greater than .7) can affect emotions and attention, and also evaluations (e.g., Zhang et al. 2022; Zhou et al. 2021), so we control for them using pixel-level color features.

*Aspects of the Presentation*. Aspects of the presentation itself may also play a role. If a video is viewed by a larger audience, for example, it may receive more likes, so we control for the (log) number of views. Similarly, if a video is available in multiple language subtitles, it may reach a broader audience, so we control for the number of available language subtitles. Certain categories (e.g., comedy rather than gaming) might be of broader interest, and more likely to be liked, so we included dummies for categories (Table WA3). Certain times of year, or specific annual events may also get more likes, so we control for year, month, and day fixed effects. Videos that have been available longer may also receive more likes, so we control for the time difference (in days) between when the video was published and when the metadata (e.g., likes) was scraped.

**Table WA1**: Summary Statistics

|  |  |  |
| --- | --- | --- |
| **Variable** | **Mean** | **SD** |
| Likes | 15,274 | 66,270 |
| Hand Movement  *Speaker Controls* | .50 | .27 |
| Gender  Age  Skin Tone  *Content Controls* | .33  42.39  1.95 | .47  10.05  .23 |
| Emotionality | 4.30 | .62 |
| Concreteness | 324.29 | 8.11 |
| Narrativity | 13.67 | 28.00 |
| Focus Past | 4.10 | 1.89 |
| Questions | .50 | .39 |
| Positivity | 44.94 | 18.30 |
| Word Familiarity | 591.17 | 6.70 |
| Swear Words | .02 | .06 |
| Function Words | 59.22 | 3.22 |
| Analytic | 42.89 | 17.96 |
| Clout | 67.16 | 19.58 |
| Authentic | 55.04 | 19.02 |
| Cognition | 13.10 | 2.37 |
| Drives | 5.26 | 1.81 |
| Time | 4.18 | 1.16 |
| Perception | 10.16 | 2.20 |
| Social | 12.17 | 3.34 |
| Conversation  Speaking Rate  Vocal Pitch  Vocal Pitch SD  Vocal Volume  Vocal Volume SD  HNR  Vocal Shimmer  Anger  Disgust  Fear  Happiness  Neutral  Sadness  Surprise  *Video Controls*  Duration  Motion Area  Motion Magnitude  Motion Direction  Average Scene Length  Warm Hue Proportion  Saturation  Brightness  Contrast of Brightness  Clarity  *Presentation Controls*  Views  Language Subtitles | .21  163.42  181.08  60.74  56.49  32.93  9.71  .10  8.22  6.02  3.04  17.49  49.17  7.31  8.75  13.76  9,141  3.28  3.04  10.66  .53  .47  .33  .21  .17  841,545  28.40 | .31  23.71  41.86  16.27  10.11  16.86  1.95  .10  6.65  4.80  2.63  12.05  14.74  5.65  6.38  5.38  7,125  1.26  .10  10.64  .19  .15  .12  .04  .12  2.960,061  5.38 |

*Notes*: Occupation, topics, category, and time effects are not included.

**Table WA2**: Distribution of Occupation

|  |  |  |
| --- | --- | --- |
| **Occupation** | **Frequency** | **Percentage** |
| Activist  Author  Designer  Inventor  Marketer  Other  Performer  Policy Person  Scientist  Technologist | 145  353  172  43  220  308  208  68  578  89 | 6.64  16.16  7.88  1.97  10.07  14.10  9.52  3.11  26.47  4.08 |

**Table WA3**: Distribution of Presentation Category

|  |  |  |
| --- | --- | --- |
| **Category** | **Frequency** | **Percentage** |
| Autos & Vehicles  Comedy  Education  Entertainment  Film & Animation  Gaming  Howto & Style  Music  News & Politics  Nonprofits & Activism  People & Blogs  Pets & Animals  Science & Technology  Sports  Travel & Events | 12  13  99  124  18  6  48  34  184  237  348  29  992  15  25 | .55  .60  4.53  5.68  .82  .27  2.20  1.56  8.42  10.85  15.93  1.33  45.42  .69  1.14 |

**Robustness**

*Measurement*. One could wonder whether the results are driven by the specific movement measure used. Using total movement (i.e., the sum across the frames), rather than the average, however, finds the same results (*b* = .018, SE = .009, *t* = 2.01, *p* = .044; Table WA4, Model 3). Similarly, one could argue that a frame-by-frame analysis includes minimal movements that are hardly perceptible to viewers, but an analysis every 100 frames yields the same results (*b* = .018, SE = .008, *t* = 2.28, *p* = .023; Table WA4, Model 4).

*Model*. One could wonder whether the results are somehow driven by the modeling approach used. To test this possibility, we used a Lasso penalized linear regression that eliminates non-essential controls and accounts for collinearity and potential overfitting (Tibshirani 1996). Results remain the same (*b* = .027, SE = .007, *t* = 3.66, *p* < .001; Table WA4, Model 5). Similarly, given the dependent variable is a count, one might wonder whether a count model could be more appropriate. Using a negative binominal regression (the outcome variable is overdispersed; *p* < .001, likelihood ratio test) finds the same results (*b* = .030, SE = .007, *t* = 4.21, *p* < .001; Table WA4, Model 6). Alternatively, one could wonder whether the results are driven by the particular dependent variable used (i.e., likes number). Using liking rate (likes number / views number; *b* = .0005, SE = .0001, *t* = 5.10, *p* < .001; Table WA4, Model 7), [[21]](#footnote-22) however, finds similar results. Finally, while we already controlled for some speaker aspects (i.e., occupation and gender) one could argue there are other speaker aspects, including unobservables, that might increase responses. Perhaps the number of likes a speaker receives on a talk is correlated with the number of likes they received on their earlier talks (potentially due to speaker-specific style or other idiosyncrasies), so to account for this potential serial correlation, we included cluster standard errors at the speaker level. Results are the same (*b* = .028, SE = .008, *t* = 3.57, *p* < .001; Table WA4, Model 8).[[22]](#footnote-23)

**Table WA4**: Main Model and Robustness Tests

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **(1)** | **(2)** | | | **(3)** | **(4)** | | **(5)** | **(6)** | **(7)** | **(8)** |
| **Hand Movement**  Controls  *Speaker*  Occupation  Gender  Age  Skin Tone  *Content*  Topics  Emotionality  Concreteness  Narrativity  Focus Past  Questions  Positivity  Word Familiarity  Swear Words  Function Words  Analytic  Clout  Authentic  Cognition  Drives  Time  Perception  Social  Conversation  Speaking Rate  Vocal Pitch  Vocal Pitch SD  Vocal Volume  Vocal Volume SD  HNR  Vocal Shimmer  Anger  Disgust  Fear  Happiness  Neutral  Sadness  *Video*  Duration  Motion Area  Motion Magnitude  Motion Direction  Average Scene Length  Warm Hue Proportion  Saturation  Brightness  Contrast of Brightness  Clarity  *Presentation*  Views  Language Subtitles  Category  Time Fixed Effects  Time Difference  N | **.028\*\*\*(.008)**  Included  -.006\*\*\*(.016)  -.011\*\*\*(.008)  .014\*\*\*(.008)  Included  .009\*\*\*(.010)  .035\*\*\*(.016)  .004\*\*\*(.007)  .005\*\*\*(.008)  .002\*\*\*(.009)  -.019\*\*\*(.009)  .004\*\*\*(.016)  .014\*\*\*(.008)  .014\*\*\*(.018)  -.003\*\*\*(.015)  -.027\*\*\*(.016)  -.004\*\*\*(.014)  .012\*\*\*(.012)  .010\*\*\*(.011)  -.017\*\*\*(.010)  -.030\*\*\*(.013)  -.029\*\*\*(.019)  -.002\*\*\*(.009)  .027\*\*\*(.010)  -.007\*\*\*(.012)  -.007\*\*\*(.010)  .003\*\*\*(.014)  -.002\*\*\*(.015)  -.006\*\*\*(.012)  .012\*\*\*(.011)  .013\*\*\*(.011)  .002\*\*\*(.010)  -.009\*\*\*(.009)  .004\*\*\*(.016)  .018\*\*\*(.019)  .014\*\*\*(.010)  .074\*\*\*(.011)  .038\*\*\*(.012)  .004\*\*\*(.009)  .009\*\*\*(.008)  .001\*\*\*(.008)  .008\*\*\*(.008)  .001\*\*\*(.010)  -.070\*\*\*(.020)  -.007\*\*\*(.010)  .068\*\*\*(.021)  1.013\*\*\*(.007)  .043\*\*\*(.012)  Included  Included  Included  2,184 | | **.069\*\*\*(.018)**  Included  -.037\*\*\*(.037)  -.060\*\*\*(.019) .055\*\*\*(.018)  Included  .054\*\*\*(.022)  .164\*\*\*(.037)  .017\*\*\*(.017)  .032\*\*\*(.018)  -.038\*\*\*(.020)  .172\*\*\*(.021)  -.046\*\*\*(.037)  .028\*\*\*(.018)  .107\*\*\*(.042)  -.033\*\*\*(.034)  -.013\*\*\*(.037)  .051\*\*\*(.031)  -.047\*\*\*(.027)  .040\*\*\*(.025)  -.053\*\*\*(.023)  -.003\*\*\*(.030)  -.035\*\*\*(.044)  .015\*\*\*(.021)  -.028\*\*\*(.023)  -.010\*\*\*(.027)  .009\*\*\*(.022)  .053\*\*\*(.035)  .039\*\*\*(.036)  .008\*\*\*(.027)  .064\*\*\*(.025)  .011\*\*\*(.025)  .037\*\*\*(.021)  .002\*\*\*(.020)  .004\*\*\*(.036)  .004\*\*\*(.043)  .005\*\*\*(.023)  .072\*\*\*(.025)  .096\*\*\*(.028)  .015\*\*\*(.021)  .034\*\*\*(.019)  -.021\*\*\*(.019)  -.022\*\*\*(.019)  -.004\*\*\*(.023)  -.007\*\*\*(.046)  .003\*\*\*(.022)  .059\*\*\*(.048)  .177\*\*\*(.016)  -.003\*\*\*(.029)  Included  Included  Included  2,160 | **.018\*\*\*(.009)**  Included  -.008\*\*\*(.016)  -.011\*\*\*(.008)  .014\*\*\*(.008)  Included  .009\*\*\*(.010)  .035\*\*\*(.016)  .003\*\*\*(.007)  .005\*\*\*(.008)  .002\*\*\*(.009)  -.018\*\*\*(.009)  .006\*\*\*(.016)  .014\*\*\*(.008)  .012\*\*\*(.018)  -.004\*\*\*(.015)  -.025\*\*\*(.016)  -.004\*\*\*(.014)  .012\*\*\*(.012)  .009\*\*\*(.011)  -.017\*\*\*(.010)  -.031\*\*\*(.013)  -.031\*\*\*(.019)  -.003\*\*\*(.009)  .027\*\*\*(.010)  -.006\*\*\*(.012)  -.007\*\*\*(.010)  .003\*\*\*(.014)  -.002\*\*\*(.016)  -.006\*\*\*(.012)  .012\*\*\*(.011)  .013\*\*\*(.011)  .003\*\*\*(.010)  -.008\*\*\*(.009)  .006\*\*\*(.016)  .018\*\*\*(.019)  .014\*\*\*(.010)  .068\*\*\*(.011)  .039\*\*\*(.012)  .005\*\*\*(.009)  .009\*\*\*(.008)  -.001\*\*\*(.008)  .008\*\*\*(.008)  .001\*\*\*(.010)  -.070\*\*\*(.020)  -.007\*\*\*(.010)  .070\*\*\*(.021)  1.014\*\*\*(.007)  .044\*\*\*(.012)  Included  Included  Included  2,184 | | **.018\*\*\*(.008)**  Included  -.007\*\*\*(.016)  -.011\*\*\*(.008)  .014\*\*\*(.008)  Included  .009\*\*\*(.010)  .036\*\*\*(.016)  .004\*\*\*(.007)  .004\*\*\*(.008)  .002\*\*\*(.009)  -.018\*\*\*(.009)  .006\*\*\*(.016)  .015\*\*\*(.008)  .014\*\*\*(.018)  -.001\*\*\*(.014)  -.026\*\*\*(.016)  -.003\*\*\*(.014)  .013\*\*\*(.012)  .010\*\*\*(.011)  -.018\*\*\*(.010)  -.032\*\*\*(.013)  -.030\*\*\*(.019)  -.003\*\*\*(.009)  .027\*\*\*(.010)  -.006\*\*\*(.011)  -.007\*\*\*(.010)  .002\*\*\*(.014)  -.003\*\*\*(.015)  -.006\*\*\*(.012)  .012\*\*\*(.011)  .012\*\*\*(.011)  .002\*\*\*(.011)  -.008\*\*\*(.009)  .005\*\*\*(.016)  .019\*\*\*(.019)  .013\*\*\*(.010)  .073\*\*\*(.011)  .037\*\*\*(.012)  .002\*\*\*(.010)  .009\*\*\*(.008)  .003\*\*\*(.008)  .008\*\*\*(.008)  -.002\*\*\*(.010)  -.068\*\*\*(.020)  -.006\*\*\*(.010)  .067\*\*\*(.021)  1.013\*\*\*(.007)  .045\*\*\*(.012)  Included  Included  Included  2,184 | **.027\*\*\*(.007)**  Included  .013\*\*\*(.007)  Included  .007\*\*\*(.009)  .020\*\*\*(.011)  \*\*\*  \*\*\*  \*\*\*  -.023\*\*\*(.008)  \*\*\*  .012\*\*\*(.008)  \*\*  -.038\*\*\*(.008)  \*\*\*  \*\*\*  -.020\*\*\*(.008)  -.031\*\*\*(.009)  .030\*\*\*(.008)  -.001\*\*\*(.011)  -.010\*\*\*(.009)  -.005\*\*\*(.009)  -.008\*\*\*(.011)  .014\*\*\*(.010)  .003\*\*\*(.008)  -.014\*\*\*(.007)  -.010\*\*\*(.008)  \*\*\*\*  .067\*\*\*(.010)  .027\*\*\*(.010)  \*\*\*  .007\*\*\*(.007)  \*\*\*  -.008\*\*\*(.008)  1.012\*\*\*(.006)  .047\*\*\*(.011)  Included  Included  Included  2,184 | | **.030\*\*\*(.007)**  Included  -.004\*\*\*(.015)  -.006\*\*\*(.008)  .013\*\*\*(.007)  Included  .010\*\*\*(.009)  .029\*\*\*(.015)  .002\*\*\*(.007)  .005\*\*\*(.007)  -.002\*\*\*(.008)  -.018\*\*\*(.009)  -.002\*\*\*(.015)  .020\*\*\*(.007)  .025\*\*\*(.017)  .005\*\*\*(.013)  -.027\*\*\*(.015)  -.003\*\*\*(.013)  .009\*\*\*(.011)  .010\*\*\*(.010)  -.016\*\*\*(.009)  -.026\*\*\*(.012)  -.025\*\*\*(.018)  .001\*\*\*(.008)  .023\*\*\*(.009)  -.003\*\*\*(.008)  -.008\*\*\*(.008)  .005\*\*\*(.008)  .004\*\*\*(.008)  -.008\*\*\*(.008)  .011\*\*\*(.008)  .010\*\*\*(.010)  .001\*\*\*(.009)  -.011\*\*\*(.008)  .007\*\*\*(.015)  .020\*\*\*(.017)  .015\*\*\*(.009)  .066\*\*\*(.010)  .035\*\*\*(.011)  .003\*\*\*(.009)  .010\*\*\*(.008)  -.001\*\*\*(.008)  .007\*\*\*(.008)  -.002\*\*\*(.009)  -.060\*\*\*(.018)  -.003\*\*\*(.009)  .059\*\*\*(.019)  1.02\*\*\*(.007)  .034\*\*\*(.012)  Included  Included  Included  2,184 | **.0005\*\*\*(.000)**  Included  -.0001\*\*\*(.000)  -.0002\*\*\*(.000)  .0002\*\*\*(.000)  Included  .0002\*\*\*(.000)  .0005\*\*\*(.000)  .0000\*\*\*(.000)  .0001\*\*\*(.000)  .0000\*\*\*(.000)  -.0002\*\*\*(.000)  -.0000\*\*\*(.000)  .0003\*\*\*(.000)  .0003\*\*\*(.000)  .0000\*\*\*(.000)  -.0004\*\*\*(.000)  .0001\*\*\*(.000)  -.0000\*\*\*(.000)  .0001\*\*\*(.000)  -.0003\*\*\*(.000)  -.0004\*\*\*(.000)  -.0003\*\*\*(.000)  .0000\*\*\*(.000)  .0004\*\*\*(.000)  -.0002\*\*\*(.000)  -.0000\*\*\*(.000)  -.0000\*\*\*(.000)  -.0000\*\*\*(.000)  -.0002\*\*\*(.000)  -.0001\*\*\*(.000)  .0002\*\*\*(.000)  .0000\*\*\*(.000)  -.0001\*\*\*(.000)  .0001\*\*\*(.000)  .0004\*\*\*(.000)  .0003\*\*\*(.000)  .0011\*\*\*(.000)  .0004\*\*\*(.000)  .0000\*\*\*(.000)  .0001\*\*\*(.000)  -.0001\*\*\*(.000)  .0001\*\*\*(.000)  .0000\*\*\*(.000)  -.0009\*\*\*(.000)  -.0001\*\*\*(.000)  .0009\*\*\*(.000)  -  .0008\*\*\*(.000)  Included  Included  Included  2,184 | **.028\*\*\*(.008)**  Included  -.006\*\*\*(.016)  -.012\*\*\*(.008)  .014\*\*\*(.007)  Included  .009\*\*\*(.009)  .035\*\*\*(.016)  .004\*\*\*(.007)  .005\*\*\*(.008)  .002\*\*\*(.010)  -.019\*\*\*(.009)  .000\*\*\*(.017)  .014\*\*\*(.009)  .014\*\*\*(.018)  -.003\*\*\*(.015)  -.027\*\*\*(.016)  -.004\*\*\*(.014)  .012\*\*\*(.012)  .010\*\*\*(.011)  -.017\*\*\*(.010)  -.030\*\*\*(.014)  -.029\*\*\*(.019)  -.002\*\*\*(.010)  .027\*\*\*(.010)  -.007\*\*\*(.011)  -.007\*\*\*(.009)  .003\*\*\*(.013)  -.002\*\*\*(.016)  -.006\*\*\*(.011)  .012\*\*\*(.011)  .013\*\*\*(.011)  .002\*\*\*(.009)  -.009\*\*\*(.009)  .004\*\*\*(.017)  .018\*\*\*(.019)  .014\*\*\*(.010)  .074\*\*\*(.013)  .038\*\*\*(.013)  .004\*\*\*(.010)  .009\*\*\*(.008)  .001\*\*\*(.007)  .008\*\*\*(.008)  .001\*\*\*(.010)  -.070\*\*\*(.020)  -.007\*\*\*(.010)  .069\*\*\*(.021)  1.013\*\*\*(.009)  .044\*\*\*(.014)  Included  Included  Included  2,184 |

*Notes*: \*\*\**p* < .001, \*\**p* < .01, \**p* < .05. Standard errors are in parentheses. **Model 1**: Main model. **Model 2**: Tone positivity in consumers’ comments as DV. **Model 3**: Total movement as focal independent variable. **Model 4**: Average movement every 100 frames as focal independent variable. **Model 5**: Lasso penalized regression. **Model 6**: Negative binomial regression. **Model 7:** Liking rate as DV. **Model 8**: With cluster standard errors at the speaker level.

**Study 2**

**Instructions for Gemini**

“You are an expert in analyzing human gestures and giving public speeches. Please help me analyze and classify peopl’'s gestures when giving speeches.

Gestures are hand movements that have a precise communication goal. They aim to carry meaning and information that complements or supplements the speech.  There are two different types of gestures: “Illustrators” and “Highlighters”, and each of them has a specific communication goal. There’s also a class named “Unrelated Movement”, which occurs when the speaker moves their hands while speaking, but such movement is irrelevant or unrelated to the spoken content (i.e., it has no communicative goal). Finally, there are cases in which the communicator is framed but they do not move their hands while speaking.

**Illustrators** {Communicative Goal: Visually depict, mimic, or resemble concrete objects (i.e., existing in the real world), actions involving concrete objects that are being spoken about, abstract concepts, or metaphorical expressions. They represent something that was happening, literally or metaphorically, but is not physically there now (i.e., it’s not physically present while the speaker is talking). Characteristics: They coincide with a part of the spoken utterance. Importantly, the key here is the object they are talking about should not be physically present at the time they are talking. Examples include: drinking gesture while saying “I was drinking water”, resembling the shape of the stairs while saying “the steps of my apartment”, throwing gesture while saying “I threw the soccer ball”, putting the hand in a gun position while saying “they shoot”, making a square shape with hands while saying “thinking outside the box”, resembling the shape of the stairs while saying “the steps of success”, throwing gesture while saying “I threw my life away”, mimicking dropping a card on the table while saying “play is our adaptive wildcard in the evolutionary process”, resembling a gap between hands while saying “the gap between doing something and doing nothing”}.

**Highlighters** {Communicative Goal: These gestures serve to underline and emphasize a spoken message without visually depicting an action, object, or abstract concept. They solely intensify the speake’'s spoken words, adding a layer of emphasis that reinforces the messag’'s urgency, importance, or emotional resonance. Examples include clapping hands together while saying “Boom. No ifs, ands or buts”, putting hands on the head while saying “Oh My God”, holding up three fingers while saying “cooperation, creativity, equality”. Attention: Most of the times, gestures that coincide with the spoken text are Illustrators. However, when the gesture coincides with the spoken text but it is not visually depicting the subject discussed, but just emphasizing part of the message, it is a Highlighter. An example includes a gesture of holding up two fingers while saying “there were two ideas”. In such case, the “two fingers” gesture coincides with the spoken content, but it’s not visually depicting the subject discussed (i.e., the ideas), but only emphasizing part of the message. In this case it should be labeled as Highlighter. The communicative goal of Highlighters also extends to when the speaker points to, or draws attention to, something that is physically present while they are talking. In this case, it is key here that if the subject discussed is physically present at the time the speaker is talking, the gesture is classified as Highlighter. Examples include: holding up two cell phones while saying “I have two cell phones”, pointing to the slides while saying “this photo here,” raising a water bottle while saying “look at this bottle,” pointing to themselves while saying “this is what I feel,” pointing to the audience while saying “you are responsible for the environment”}.

**Unrelated Movements** {When the speaker moves their hands while speaking, but such movement is irrelevant or unrelated to the spoken content (i.e., it has no communicative goal), then it should be labeled as Unrelated Movement. Unrelated Movement is hand movement accompanying the speech, but not carrying any informational content. When you identify hand movements, but it’s not Illustrator or Highlighter, please label it as Unrelated Movement}.

**No Movement** {When the speaker is framed but does not move their hands}.

In this analysis task, you will examine a sequence of elements from each video example organized in a specific format. The first element in each request is the transcript of a speech given by a speaker. Following the transcript, the request includes a comprehensive series of frame data extracted from the entire video of the speech, rather than only the first few frames. Each piece of frame data is structured to contain two key components: (1) the timestamp indicating when during the video the frame was captured (e.g., 0.10s, 0.20s). These timestamps are crucial as they help indicate the precise moment each frame was captured, corresponding to 10 frames per second, and (2) a reference to the image file representing that frame. You should focus on analyzing the gestures in all these images throughout the entire video and their relationship with the speech text, which is very important. It is vital to consider the temporal progression of the gestures as demonstrated across the full sequence of frames for accurate classification. Please classify each example into one of the four categories. Specifically, only one result for each set of sequential images. Output the results in the following JSON format:

{“class”: “Illustrator or Highlighter or Unrelated Movement or No Movement”,

“reasoning”: “briefly describe your reasoning behind the classification”,

“hand movement: “yes or no”}.”

**Solving LMM’S Classification Ambiguities**

On the initial pass, Gemini occasionally misclassified gestures due to challenges in distinguishing between categories. Based on these misclassifications, the instructions were fine-tuned multiple times. Main sources of initial misclassifications included:

* Highlighters Misclassified as Illustrators. This primarily occurred in two situations:

1. When the physical object was present. For example, when a speaker said, “Look at this!” while pointing at a physically present object, Gemini initially classified it as an Illustrator. Based on prior theoretical work (e.g., McNeil 1992), however, drawing attention to something physically present (e.g., through pointing), does not involve using the hands to visually depict it. Therefore, the gesture should be classified as a Highlighter.
2. When speech coincided with the gesture. For example, a speaker holding up three fingers while saying, “There are three houses.” In this instance, the “three fingers” gesture aligns with the spoken content but does not visually depict the subject discussed (i.e., the houses). Instead, the gesture emphasizes part of the message, making it a Highlighter.

* Illustrators Misclassified as Highlighters**.** This mainly occurred when the illustrator was metaphoric – visually depicting an abstract or metaphorical concept (e.g., “the steps of success”) rather than a concrete object or action (e.g., “the steps of the house”). When a speaker said “contribute and collaborate” while bringing their hands together, for example, Gemini initially classified the gesture as a Highlighter, interpreting it as a movement emphasizing the concept. However, the gesture symbolically represents the concept of coming together of efforts or resources, which is central to the metaphorical idea of contribution and collaboration, making it an Illustrator.
* Illustrators or Highlighters Misclassified as Unrelated Movement. This occurred in two situations:

1. When the hand movements were subtle. Subtle movements were difficult for Gemini to detect and interpret, leading to their classification as Unrelated Movements, while they were Illustrators or Highlighters, depending on the speech.
2. When the physical object was present, but Gemini could not “see” it. For example, when a speaker said “That person” while pointing at a physically present object (i.e., individual), Gemini initially classified the gesture as an Unrelated Movement, rather than a Highlighter, because it could not detect the object being referred to.

**Matching RAs with Gemini’s Classification**

Two research assistants coded a set of 130 TED talks, in terms of number of illustrators and highlighters. Along with the gesture type, they reported the time of the video in which the gesture occurred (e.g., 2:34-2:36: Illustrator). To also obtain the number of unrelated movements and no movements at the video level, we matched the RAs’ coding with Gemini’s one (based on 10-seconds intervals) based on the rules in Table WA6.

**Table WA6**: Matching Rules

|  |  |  |  |
| --- | --- | --- | --- |
| **Time of the Video** | **Gemini Classification** | **RAs’ Classification** | **Final Classification** |
| 00:01-00:10 | Highlighter | Highlighter | Highlighter |
| 00:11-00:20 | Illustrator | Illustrator | Illustrator |
| 00:21-00:30 | Highlighter/ Unrelated /No Movement | Illustrator | Illustrator |
| 00:31-00:40 | Illustrator/ Unrelated /No Movement | Highlighter | Highlighter |
| 00:41-00:50 | No Movement | Nothing | No Movement |
| 00:51-01:00 | Illustrator/Highlighter/ Unrelated | Nothing | Unrelated |

**Web Appendix B**

**Study 3a**

**Exclusions, Demographics, and Experimental Material**

Following the preregistration, we sought to collect 200 participants per condition from Prolific. However, due to technical issues encountered by many participants when playing the video, we ended up collecting 1,004.[[23]](#footnote-24) As pre-registered, we excluded participants who failed any of the three attention checks: “The speaker was: man, woman” (N = 9), “The product pitched was: water system filtration, language learning app, skincare product” (N = 3), or “How much did you enjoy taking this survey? Please indicate 2” (N = 11). This resulted in a final sample of 685 (54.5% women, 42.8% men, .7% other; mean age = 40.1 years).

The speech read, “Want to revitalize your skin? RadianceGlow is a natural cream. Just spread it on your face, massaging gently in circular motion every morning and night. Your skin will stay smooth and vibrant all the time.”

**Dependent Variables**

*Speaker Liking*. An ANOVA revealed a significant effect of movement type on speaker liking (F(3, 681) = 3.03, *p* = .029). Specifically, pairwise comparisons found that using illustrators (M = 4.77) made observers like the speaker more than highlighters (M = 4.38, *b* = .38, SE = .14, *t* = 2.72, *p* = .007), unrelated movements (M = 4.46, *b* = .32, SE = .14, *t* = 2.34, *p* = .020) or no movement at all (M = 4.45, *b* = .31, SE = .14, *t* = 2.15, *p* = .032).

*Product Liking*. An ANOVA revealed a significant effect of movement type on product liking (F(3, 681) = 2.69, *p* = .045). Specifically, pairwise comparisons found that using illustrators (M = 4.11) made observers like the product more than highlighters (M = 3.69, *b* = .41, SE = .16, *t* = 2.58, *p* = .010), unrelated movements (M = 3.76, *b* = .32, SE = .15, *t* = 2.09, *p* = .037) or no movement at all (M = 3.78, *b* = .34, SE = .16, *t* = 2.12, *p* = .034).

*Purchase Intention*. An ANOVA revealed a significant effect of movement type on purchase intention (F(3, 681) = 2.69, *p* = .045). Specifically, pairwise comparisons found that using illustrators (M = 3.87) made observers more interested in buying the product than highlighters (M = 3.40, *b* = .47, SE = .18, *t* = 2.57, *p* = .010), unrelated movements (M = 3.54, *b* = .50, SE = .18, *t* = 2.82, *p* = .005) or no movement at all (albeit marginally, M = 3.37, *b* = .33, SE = .19, *t* = 1.78, *p* = .075).

**Other Mediation Results**

*Mediation (Illustrators vs. Unrelated)*. Serial mediation analysis with 5,000 bootstrap resamples (Preacher and Hayes 2004) found that understandability and perceived competence serially mediated illustrator’s effects on persuasion (indirect effect = .08, 95% CI = .02, .14). Illustrators made the content easier to understand (*b* = .57, SE = .12, *t* = 4.90, *p* < .001), which made the speaker seem more competent (*b* = .50, SE = .04, *t* = 10.92, *p* < .001), which boosted persuasion (*b* = .73, SE = .05, *t* = 14.77, *p* < .001). Including these two mediators led the conditional direct effect to be reduced to non-significance (95% CI = –.21, .26).

*Mediation (Illustrators vs. No Movement)*. Serial mediation analysis with 5,000 bootstrap resamples (Preacher and Hayes 2004) found that understandability and perceived competence serially mediated illustrator’s effects on persuasion (indirect effect = .21, 95% CI = .11, .30). Illustrators made the content easier to understand (*b* = .34, SE = .11, *t* = 2.27, *p* = .003), which made the speaker seem more competent (*b* = .53, SE = .05, *t* = 10.46, *p* < .001), which boosted persuasion (*b* = .70, SE = .05, *t* = 13.07, *p* < .001). Including these two mediators led the conditional direct effect to be reduced to non-significance (95% CI = –.19, .26).

**Ancillary Analyses**

*Vividness*. As expected, an ANOVA revealed a significant effect of movement type on vividness (F(3, 681) = 12.91, *p* < .001). Specifically, pairwise comparisons found that using illustrators (M = 4.91) made the content more vivid, concrete, and tangible than using highlighters (M = 4.23, *b* = .68, SE = .16, *t* = 4.18, *p* < .001), unrelated movements (M = 3.91, *b* = .56, SE = .16, *t* = 3.52, *p* < .001) or no movement at all (M = 4.36, *b* = 1.00, SE = .16, *t* = 6.05, *p* < .001).

*Audience Expertise.* Audience expertise did not appear to moderate illustrator’s effects on understandability compared to highlighters (F = .14, *p* = .713), unrelated movements (F = 1.20, *p* = .275), or no movement (F = 1.76, *p* = .185). Given expertise might be a function on both gender (e.g., women could be more expert in skincare products than men) and age (e.g., older consumers might have more knowledge), this model controls for gender and age. Results are similar without controls.

**Study 3b**

**Method**

Following the preregistration, we sought to collect 900 participants pre-exclusion from Prolific. However, due to technical issues some participants encountered when playing the video (N = 127), we ended up collecting a total of 1,085.[[24]](#footnote-25) As pre-registered, we excluded duplicate IP addresses (N = 21) and participants who failed at least one of the two attention checks: “The speaker was: man, woman” (N = 35) or “The product pitched was: water system filtration, language learning app, skincare product” (N = 32). This resulted in a final sample of 892 (56.1% women, 42.3% men, 1.6% other; mean age = 38.1 years).

The procedure was similar to Study 3a.Participants were randomly assigned to one of four conditions (movement type: illustrator vs. highlighter vs. unrelated vs. no movement) in which they watched a short video of an entrepreneur pitching a new product. To assess robustness (i.e., stimulus-sampling; Wells and Windschitl 1999), we tested whether the results held across three different pitches (i.e., language learning app, water filtration system, skincare cream; one condition per participant, in a between-subjects design). As in Study 3a, the only difference between conditions was the type of hand movement (the words were identical).[[25]](#footnote-26) In addition, to test robustness and generalizability, the exact gesture used in each recording varied across the products (e.g., the illustrator used in the skincare pitch was different from the one used in the language app pitch). This helps cast doubt on the notion that any results are driven by the specific gesture used in a given pitch.

After watching the video, participants completed the dependent variables. In addition to the persuasion measure (α = .88) from Study 3a,[[26]](#footnote-27) participants were asked how likely they were to recommend the product to others (1 = very unlikely to recommend, 7 = very likely to recommend).

Next, we measured potential underlying processes. Understandability was the same of Study 3a and given that one’s confidence in a topic is a key dimension of competence (e.g., Fiske et al. 2002; Kervyn, Fiske, and Malone 2012; Wang et al. 2021), to test robustness to an alternate measure of competence, participants were asked to rate how confident they thought the speaker was with what they were saying (1 = not at all confident, 7 = extremely confident). Participants also completed measures of caring, extraversion, genuineness, attention, and mental simulation, a measure of movement engagingness, [[27]](#footnote-28) as well as an alternative explanation based on movement typicality (see below for an explanation behind each and the measures used). To ensure that the order in which participants completed the various measures did not affect their responses, item order was randomized.

Finally, participants completed the attention checks and demographics.

**Results**

*Persuasion*. Consistent with Studies 2 and 3a, an ANOVA revealed a significant effect of movement type on persuasion (F(3, 888) = 3.15, *p* = .024). Specifically, pairwise comparisons found that using illustrators (M = 4.92) increased persuasion more than using highlighters (M = 4.60, *b* = .32, SE = .13, *t* = 2.38, *p* = .017), unrelated movements (M = 4.60, *b* = .32, SE = .14, *t* = 2.32, *p* = .020) or no movement at all (M = 4.56, *b* = .36, SE = .13, *t* = 2.73, *p* = .006). There was no difference between conditions on willingness to recommend (F(3, 888) = .91, *p* = .437).[[28]](#footnote-29)

*Understandability.* Using illustrators also increased understandability (albeit marginally, MIllustrator = 6.18 vs. MHighlighter = 5.96, *b* = .22, SE = .11, *t* = 1.08, *p* = .060; MIllustrator = 6.18 vs. MUnrelated = 5.94, *b* = .24, SE = .12, *t* = 2.07, *p* = .039; MIllustrator = 6.18 vs. MNo movement = 4.94, *b* = .24, SE = .11, *t* = 2.07, *p* = .039)

*Competence.* It also made the speaker seem more competent (MIllustrator = 5.22 vs. MHighlighter = 4.87, *b* = .35, SE = .16, *t* = 2.18, *p* = .029; MIllustrator = 5.22 vs. MUnrelated = 4.71, *b* = .50, SE = .16, *t* = 3.07, *p* = .002; MIllustrator = 5.22 vs. MNo movement = 4.77, *b* = .44, SE = .16, *t* = 2.80, *p* = .005).

*Mediation (Illustrators vs. Highlighters)*. Serial mediation analysis with 5,000 bootstrap resamples (Preacher and Hayes 2004) found that understandability and perceived competence serially mediated illustrator’s effects on persuasion (albeit marginally, *p* = .065, indirect effect = .05, 95% CI = –.003, .11). Using illustrators increased understandability (*b* = .22, SE = .11, *t* = 1.96, *p* = .050), which made the speaker seem more competent (*b* = .52, SE = .07, *t* = 7.41, *p* < .001), which boosted persuasion (*b* = .47, SE = .03, *t* = 14.49, *p* < .001). The conditional direct effects to be reduced to non-significance (95% CI = –.08, .29).

*Mediation (Illustrators vs. Unrelate Movements)*. Understandability and perceived competence serially mediated illustrator’s effects on persuasion (albeit marginally, *p* = .051, indirect effect = .05, 95% CI = –.0003, .11). Using illustrators increased understandability (*b* = .22, SE = .11, *t* = 1.96, *p* = .050), which made the speaker seem more competent (*b* = .52, SE = .07, *t* = 7.41, *p* < .001), which boosted persuasion (*b* = .45, SE = .03, *t* = 12.95, *p* < .001). The conditional direct effect was not significant (95% CI = –.18, .22).

*Mediation (Illustrators vs. No Movement)*. Understandability and perceived competence serially mediated illustrator’s effects on persuasion (indirect effect = .05, 95% CI = .001, .10). Using illustrators increased understandability (*b* = .22, SE = .11, *t* = 1.96, *p* = .050), which made the speaker seem more competent (*b* = .52, SE = .07, *t* = 7.41, *p* < .001), which boosted persuasion (*b* = .46, SE = .03, *t* = 14.14, *p* < .001). The conditional direct effect was not significant (95% CI = –.08, .30).

**Results with the three measures of persuasion analyzed separately**

*Speaker Liking*. An ANOVA revealed a significant effect of movement type on speaker liking (F(3, 888) = 4.71, *p* = .003).[[29]](#footnote-30) Specifically, pairwise comparisons found that using illustrators (M = 5.21) made observers like the speaker more than highlighters (M = 4.92, *b* = .28, SE = .13, *t* = 2.20, *p* = .028), unrelated movements (M = 4.82, *b* = .38, SE = .13, *t* = 2.90, *p* = .004) or no movement at all (M = 4.76, *b* = .45, SE = .13, *t* = 3.51, *p* < .001).

*Product Liking.* Using illustrators also made participants like the product more (albeit marginally, MIllustrator = 4.85 vs. MHighlighter = 4.48, *b* = .27, SE = .15, *t* = 1.83, *p* = .067; MIllustrator = 4.85 vs. MUnrelated = 4.57, *b* = .29, SE = .15, *t* = 1.86, *p* = .064; MIllustrator = 4.85 vs. MNo movement = 4.57, *b* = .28, SE = .15, *t* = 1.91, *p* = .057)

*Purchase Intention.* Using illustrators also made participants more interested in buying the product (MIllustrator = 4.71 vs. MHighlighter = 4.31, *b* = .40, SE = .17, *t* = 2.39, *p* = .017; MIllustrator = 4.71 vs. MUnrelated = 4.42, *b* = .29, SE = .17, *t* = 1.68, *p* = .094; MIllustrator = 4.71 vs. MNo movement = 4.35, *b* = .35, SE = .16, *t* = 2.15, *p* = .032).

*Mediation (Illustrators vs. Highlighters).* Serial mediation analysis with 5,000 bootstrap resamples (Preacher and Hayes 2004) found that understandability and perceived competence serially mediated illustrator’s effects (albeit marginally, speaker liking: *p* = .069, 95% CI = –.003, .09; product liking: *p* = .064, 95% CI = –.003, .011; purchase: *p* = .066, 95% CI = –.004, .13). Using illustrators increased understandability (*b* = .22, SE = .11, *t* = 1.96, *p* = .050), which made the speaker seem more competent (*b* = .52, SE = .07, *t* = 7.41, *p* < .001), which made observers like them more (*b* = .40, SE = .03, *t* = 11.69, *p* < .001), and the product more (*b* = .48, SE = .04, *t* = 11.84, *p* < .001), and made observers more interested in buying the product (*b* = .54, SE = .04, *t* = 12.41, *p* < .001). The conditional direct effect was not significant (speaker liking: 95% CI = –.11, .27; product liking: 95% CI = –.16, .29; purchase: 95% CI = –.10, .42).

*Mediation (Illustrators vs. Unrelated Movements).* Understandability and perceived competence serially mediated illustrator’s effects (albeit marginally, speaker liking: *p* = .054, 95% CI = –.0007, .09; product liking: *p* = .052, 95% CI = –.0004, .11; purchase: 95% CI = –.0004, .12). Using illustrators increased understandability (*b* = .24, SE = .12, *t* = 2.05, *p* = .041), which made the speaker seem more competent (*b* = .49, SE = .07, *t* = 7.30, *p* < .001), which made observers like them more (*b* = .39, SE = .04, *t* = 10.47, *p* < .001), and the product more (*b* = .47, SE = .04, *t* = 10.92, *p* < .001), and made observers more interested in buying the product (*b* = .49, SE = .04, *t* = 10.98, *p* < .001). The conditional direct effect was not significant (speaker liking: 95% CI = –.09, .32; product liking: 95% CI = –.27, .22; purchase: 95% CI = –.30, .23).

*Mediation (Illustrators vs. No Movement).* Understandability and perceived competence serially mediated illustrator’s effects (speaker liking: 95% CI = .0001, .09; product liking: 95% CI = .001, .10; purchase: 95% CI = .0009, .12). Using illustrators increased understandability (*b* = .24, SE = .11, *t* = 2.15, *p* = .031), which made the speaker seem more competent (*b* = .47, SE = .06, *t* = 7.22, *p* < .001), which made observers like them more (*b* = .39, SE = .04, *t* = 10.74, *p* < .001), and the product more (*b* = .47, SE = .04, *t* = 11.93, *p* < .001), and made observers more interested in buying the product (*b* = .52 SE = .04, *t* = 12.14, *p* < .001). The conditional direct effect was not significant (speaker liking: 95% CI = –.02, .42; product liking: 95% CI = –.20, .25; purchase: 95% CI = –.19, .34).

**Alternative Processes and Other Significant Mediation Results**

Following the pre-registration, we collected measures of extraversion, perceived genuineness, perceived caring, attention, mental simulation, and movement typicality. While the other variables sometimes differed between some conditions, understandability and competence were the only factors that explained the difference between illustrators and *all* three other conditions. Below, we report the measures used, as well as any significant mediations of illustrators vis-à-vis the other movement types on persuasion. See Table WB1 for a summary.

**Table WB1**: The Effects of Illustrators

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Illustrator vs. Highlighter** | **Illustrator vs. Unrelated** | **Illustrator vs.**  **No Movement** |
| **Persuasion** | **4.92 vs. 4.60** | **4.92 vs. 4.60** | **4.92 vs. 4.56** |
| **Understandability** | 6.18 vs. 5.96 | **6.18 vs. 5.94** | **6.18 vs. 5.94** |
| **Competence** | **5.22 vs. 4.87** | **5.22 vs. 4.71** | **5.22 vs. 4.77** |
| Other Processes |  |  |  |
| Extraversion | 4.76 vs. 4.59 | **4.76 vs. 4.19** | **4.76 vs. 4.31** |
| Genuineness | 5.04 vs. 4.81 | **5.04 vs. 4.48** | **5.04 vs. 4.70** |
| Caring | **5.52 vs. 5.20** | 5.52 vs. 5.29 | 5.52 vs. 5.25 |
| Attention | 4.11 vs. 3.93 | **4.11 vs. 3.74** | **4.11 vs. 3.60** |
| Mental Simulation | **4.63 vs. 4.24** | 4.63 vs. 4.44 | **4.63 vs. 4.03** |
| *Alternative Explanation* |  |  |  |
| Typicality | 4.51 vs 4.59 | 4.51 vs 4.35 | 4.51 vs 4.24 |

*Notes*: Mean differences significant at *p* < .05 in bold.

*Extraversion (and Emotionality)*. One could wonder whether the results are driven by perceived extraversion (or emotional involvement). Using illustrators can make speakers seem more emotionally involved, engaged, and dynamic, signaling extraversion or energy. Appearing more extravert or enthusiastic, in turn, could lead audiences to like them more (Friedman, Riggio, and Casella 1988). To measure extraversion, we collected two items adapted from prior work (i.e., Wang and Ziano 2022; “how extraverted do you think the speaker is?” 1 = not at all extraverted, 7 = extremely extraverted; “how enthusiastic was the speaker when sharing their content?” 1 = not at all enthusiastic, 7 = extremely enthusiastic; *r* = .72). Pairwise comparisons found that using illustrators (M = 4.76) increased perceived extraversion more than unrelated movements (M = 4.19, *b* = .57, SE = .15, *t* = 3.73, *p* < .001) and no movement (M = 4.31, *b* = .44, SE = .15, *t* = 2.98, *p* = .003), but not compared to highlighters (M = 4.59, *b* = .16, SE = .15, *t* = 1.10, *p* = .272). Further, extraversion mediated the illustrator’s effect on persuasion compared to both unrelated (95% CI = .11, .40) and no movement (95% CI = .06, .33), but not compared to highlighters (95% CI = –.06, .20).

To further test whether emotional involvement is driving things, we tested whether there was difference across conditions on how “enthusiastic” (extraversion’s item) the speaker seemed. Results remained the same.

Additionally, we conducted a post-test (<https://aspredicted.org/q45n-p26r.pdf>) asking participants how excited, emotionally involved, and energetic the speaker seemed (α = .94) across conditions. Participants assigned (N = 77, Prolific, 52% women, 48% men, mean age = 40 years)[[30]](#footnote-31) were randomly assigned to three conditions, one different movement type for each product pitch. Further casting doubt on this alternative explanation, emotionality did not vary across conditions (F (3, 227) = .42, *p* = .742) and including participant random effects found the same results (*p*s > .56).

*Genuineness*. Alternatively, one could wonder whether the results are driven by perceived genuineness. Prior work suggests people tend to exhibit less hand movement when they lie (e.g., Burgoon, Schuetzler, and Wilson 2015). Further, given that illustrators visually reflect one’s thoughts, one could argue that they can make speakers seem more genuine (Caso et al. 2006). Perceived genuineness, in turn, can increase positive consumer evaluations (Hamilton, Vohs, and McGill 2014). To measure perceived genuineness, we collected two items of genuineness(“how genuine is the speaker? 1 = not at all genuine, 7 = very genuine; “how much do you think the speaker truly believes in what they are saying? 1 = very little 7 = very much; *r* = .90). Pairwise comparisons found that using illustrators (M = 5.04) increased perceived genuineness more than unrelated movements (M = 4.48, *b* = .36, SE = .16, *t* = 2.22, *p* = .027) and no movement (M = 4.70, *b* = .34, SE = .16, *t* = 2.14, *p* = .032), but not compared to highlighters (M = 4.81, *b* = .23, SE = .16, *t* = 1.46, *p* = .146). Further, genuineness mediated the illustrator’s effect on persuasion compared to both unrelated (95% CI = .02, .39) and no movement (95% CI = .02, .36), but not compared to highlighters (95% CI = –.04, .30).

*Caring*. One could wonder whether the results are driven by perceived caring. Given that coming up with a gesture that illustrates what one is saying requires some effort, when speakers use illustrators, audiences may infer that speakers care about helping the audience understand what they are saying. This, in turn, could foster more favorable consumer responses (Morales 2005). To measure perceived caring, participants were asked how much they thought the speaker cared about how easy it was to understand what they were saying (1 = didn’t care at all, 7 = cared a lot). Pairwise comparisons found that using illustrators (M = 5.52) increased perceived caring more than using highlighters (M = 5.20, *b* = .28, SE = .14, *t* = 2.34, *p* = .020) but only marginally compared to unrelated movements (M = 5.29, *b* = .27, SE = .14, *t* = 1.89, *p* = .059) and no movement (M = 5.25, *b* = .23, SE = .14, *t* = 1.69, *p* = .092). Further, caring mediated the illustrator’s effect on persuasion compared to highlighters (95%CI = .03, .32), but not compared to unrelated movements (95%CI = –.02, .30) or no movement at all (95%CI = –.02, .27).

*Attention*. Alternatively, one could wonder whether the results are driven by attention. Illustrators can make encourage attention to what is being said. Rather than simply stating how to apply face cream, for example, demonstrating the action (e.g., mimicking circular motions on their face) can encourage audiences to focus on the spoken words. This attention, in turn, could boost positive consumer evaluations (e.g., Pieters and Wedel 2004). To measure attention, participants were asked how attention getting they found the presentation (1 = wasn’t attention getting at all, 7 = was very attention getting). Pairwise comparisons found that using illustrators (M = 4.11) increased attention more than unrelated movements (M = 3.74, *b* = .37, SE = .18, *t* = 2.05, *p* = .041) and no movement (M = 3.60, *b* = .52, SE = .18, *t* = 2.92, *p* = .004), but not compared to highlighters (M = 3.93, *b* = .18, SE = .18, *t* = 1.02, *p* = .308). Further, attention mediated the illustrator’s effect compared to unrelated movement (95% CI = .007, .32) and no movement (95% CI = .07, .38), but not compared to highlighters (95% CI = –.08, .24).

*Mental Simulation*. Alternatively, one could wonder whether the results are driven by mental simulation. Illustrators visually depict information, which could help audiences mentally simulate product use (e.g., imagine themselves using the product). When presenting a new blender, for example, mimicking the motion of holding the blender and pressing buttons can prompt the audience to imagine themselves doing the same at home. This, in turn, could increase consumer evaluations (Elder and Krishna 2012). To measure mental simulation, participants were asked how easily they could imagine themselves using the device (1 = not easily at all, 7 = very easily; Elder and Krishna 2012). Pairwise comparisons found that using illustrators (M = 4.63) increased mental simulation more than highlighters (M = 4.24, *b* = .39, SE = .18, *t* = 2.21, *p* = .027) and no movement (M = 4.03, *b* = .60, SE = .18, *t* = 3.42, *p* = .001), but not compared to unrelated movements (M = 4.44, *b* = .19, SE = .18, *t* = 1.05, *p* = .292). Further, mental simulation mediated the illustrator’s effect on persuasion compared to both highlighters (95% CI = .02, .41) and no movement (95% CI = .14, .52), but not compared to unrelated movements (95% CI = –.09, .30).

*Movement Typicality.* Consistent with Study 3a, typicality (α = .91) did not vary across conditions (F(3, 888) = 2.47, *p* = .06).

**Additional Measure Collected (Engagingness)**

*Movement Engagingness*. We initially pre-registered movement engagingness (“how engaging was the way the speaker moved their hands?” 1 = not at all engaging, 7 = very engaging) as a potential mechanism, but upon reflection, we realized it provides limited insight into what is actually going on. Even if differences were observed across gestures, this measure would not explain why certain gestures are more engaging. Furthermore, if an illustrator is more engaging than another type of gesture, it would not be surprising if this led to more favorable impressions about the speaker (e.g., how competent they seem) and other positive outcomes.

Consistent with this notion, an ANOVA analysis found a significant effect of movement type on engagingness (F(3, 888) = 17.19, *p* < .001). Specifically, illustrators (M = 4.65) were seen as more engaging than highlighters (M = 4.24, *b* = .40, SE = .17, *t* = 2.42, *p* = .016), unrelated movements (M = 3.70, *b* = .94, SE = .17, *t* = 5.48, *p* < .001) and no movement (M = 3.58, *b* = 1.06, SE = .17, *t* = 6.38, *p* < .001). Furthermore, we tested the condition 🡪 understandability 🡪 engagingness 🡪 competence 🡪 persuasion serial path. We found significant mediation of illustrators on persuasion compared to highlighters (albeit marginal, *p* = .065, 95% CI = –.0001, .04), unrelated movements (albeit marginal, *p* = .052, 95% CI = –.0001, .03), and no movement (95% CI = .0001, .03).

**Discussion**

First, Study 3b further demonstrates the benefits of illustrators on persuasion. Using illustrators made audiences like the speaker more, boosted liking of the product discussed, and increased their interest in purchasing it.

Second, consistent with Study 3a, the effects seem to be driven, at least in part, by understandability and perceived competence. Using illustrators made the content easier to understand, which made communicators seem more competent, which boosted persuasion.

That said, we don’t mean to suggest that understandability and competence are the only factors that make illustrators beneficial. Relative to unrelated movements or no movement, for example, illustrators also increased attention. That said, the results indicate that making speakers seem more competent, by making the content easier to understand, was the *only* process that explained illustrators’ positive impact relative to *all* three other movement types.

Third, the results cast further doubt on the possibility that typicality is playing a role, and the fact that the effects hold across different product types (e.g., a language learning app and skincare cream) and gestures (e.g., different illustrators for different products) speaks to their generalizability.

**Web Appendix C**

## Study 4b: Process by Moderation (Certainty Language)

Using the field data from Study 2, Study 4b provides an additional process test. If part of the reason illustrators boost evaluations is because they make speakers seem more competent (by making the content easier to understand), then the effect should be mitigated in the presence of other cues that already signal competence. To explore this possibility, Study 4b examines the potential moderating role of certainty language (“mediation-via-moderation”; Bullock, Green, and Ha 2010). Using more certain language serves as an unambiguous, direct, and explicit marker of speaker confidence and competence in a topic (Rocklage et al. 2023).[[31]](#footnote-32) Consequently, if perceived competence is playing a role, illustrators’ positive effect should be mitigated in the presence of certain language (as there is already some evidence that the speaker is competent). To test this, we used the Certainty lexicon (Rocklage et al. 2023) to measure the level of certainty reflected by the words and phrases used,[[32]](#footnote-33) and included the illustrators × certainty language interaction term in Equation 3 (which includes all the control variables used in Study 2 previously). Certainty scores (mean = 8.35, SD = .29, ranging from 6.67 to 8.9) were normalized (mean = 0, SD = 1) to facilitate interpretation.

As predicted, results reveal a significant illustrator × certainty language interaction (*b* = –.019, SE = .007, *t* = –2.66, *p* = .008).[[33]](#footnote-34) Consistent with the underlying role of competence, illustrators’ positive effects were mitigated for higher levels of certainty language (Figure WC1).

**Figure WC1**: Moderation by Certainty Language, Floodlight Analysis

A graph showing the average marginal effects of illustrators

Description automatically generated

*Notes*: The dashed line indicates the level of certainty (Johnson-Neyman point: 0.25) at which illustrators are not beneficial anymore (right region).

To provide another interpretation of the results, we conducted a spotlight analysis at ±1SD (Figure WC2). Consistent with the underlying role of perceived competence, when speakers should have seemed less competent to begin with (i.e., because they used –1SD less certain language), talks that used more illustrators were evaluated more positively (*b* = .084, SE = .024, *t* = 3.55, *p* < .001). When speakers should have already seemed competent (i.e., because they used +1SD more certain language), though, consistent with our theorizing, illustrators’ impact was mitigated (*b* = .007, SE = .024, *t* = .30, *p* = .765). Speakers’ language should have already made them seem at least somewhat competent, and thus illustrators had less of an impact.[[34]](#footnote-35)

**Figure WC2**: Moderation by Certainty Language, Spotlight Analysis

**Discussion**

These results shed further light on a potential reason why illustrators might be beneficial. Consistent with the underlying role of perceived competence, and Studies 3a and 3b, Study 4 demonstrates that the illustrators’ positive impact was mitigated in the presence of other cues that already signaled competence (i.e., certainty-oriented language).

One could wonder whether verbal and non-verbal cues add together to shape competence perceptions. While this is certainly possible, and may hold in some cases, in this instance the effect of illustrators in high certainty language condition was nonsignificant (*p* = .765). Though it is difficult to know why, one possibility could be that the language was already certain enough in this instance, and thus the gesture did not add additional certainty (i.e., a diminishing effect).

**Reference (not in the main text)**

Beattie, Geoffrey (2004), *Visible thought: The new psychology of body language*, Routledge.

Blei, David M. (2012), “Probabilistic topic models,” *Communications of the ACM*, 55 (4), 77-84.

Boyd, Ryan L., Ashwini Ashokkumar, Sarah Seraj, and James W. Pennebaker (2022), “The development and psychometric properties of LIWC-22,” *Austin, TX: University of Texas at Austin*, *10*.

Elder, Ryan S. and Aradhna Krishna (2012), “The “visual depiction effect” in advertising: Facilitating embodied mental simulation through product orientation,” *Journal of Consumer Research*, 38 (6), 988-1003.

Friedman, Howard S., Ronald E. Riggio, and Daniel F. Casella (1988), “Nonverbal skill, personal charisma, and initial attraction,” *Personality and Social Psychology Bulletin*, 14 (1), 203-211.

Guyer, Joshua J., Leander R. Fabrigar, and Thomas I. Vaughan-Johnston (2019), “Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion,” *Personality and Social Psychology Bulletin*, *45*(3), 389-405.

Hamilton, Ryan, Kathleen D. Vohs, and Ann L. McGill (2014), “We'll be honest, this won't be the best article you'll ever read: The use of dispreferred markers in word-of-mouth communication,” *Journal of Consumer Research*, 41 (1), 197-212.

Hu, Allen, and Song Ma (2024), “Persuading investors: A video-based study,” *Journal of Finance*, forthcoming.

Kervyn, Nicolas, Susan T. Fiske, and Chris Malone (2012), “Brands as intentional agents framework: How perceived intentions and ability can map brand perception,” *Journal of Consumer Psychology*, 22 (2), 166-176.

Morales, Andrea C. (2005), “Giving firms an “E” for effort: Consumer responses to high-effort firms,” *Journal of Consumer Research*, 31 (4), 806-812.

Pieters, Rik and Michel Wedel (2004), “Attention capture and transfer in advertising: Brand, pictorial, and text-size effects,” *Journal of Marketing*, 68 (2), 36-50.

Rocklage, Matthew D., Sharlene He, Derek D. Rucker, and Loran F. Nordgren (2023), “Beyond sentiment: The value and measurement of consumer certainty in language,” *Journal of Marketing Research*, 60 (5), 870-888.

Tibshirani, Robert (1996), “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58 (1), 267-288.

Wang, Deming and Ignazio Ziano (2022), “Faster responders are perceived as more extraverted,” *Journal of Experimental Psychology: General*, 151 (12), 3177.

Zhang, Shunyuan, Dokyun Lee, Param Vir Singh, and Kannan Srinivasan (2022), “What makes a good image? Airbnb demand analytics leveraging interpretable image features,” *Management Science*, 68 (8), 5644-5666.

1. We created a custom version of ChatGPT because an equivalent Gemini tool was not available at the time of the research. For testing purposes, guidelines and a sample of video frames are available on the OSF repository. [↑](#footnote-ref-2)
2. Indeed, while it did not directly compare different types of movements empirically, Dargue, Sweller, and Jones’ (2019) meta-analysis found that all hand movements were similarly effective (i.e., hand movement type “did not significantly predict the size of the effect,” p. 777). [↑](#footnote-ref-3)
3. We computed the movement only for frames where the hands were detected, allowing for more precise measurement. We summed the movement of the left and right hand to get a unique value of movement. [↑](#footnote-ref-4)
4. While MediaPipe has been used for gesture recognition and sign language interpretation, to the best of our knowledge, this is the first time it is used to quantify hand movement magnitude over time in behavioral research. [↑](#footnote-ref-5)
5. Two research assistants (blinded to hypotheses) rated a random sample of 135 videos on how much the speaker moved their hands (1 = not at all, 5 = very much; *r* = .81). [↑](#footnote-ref-6)
6. We collected all the comments available for 2,160 talks. Using a different measure of positivity (LIWC Tone) finds the same results (*b* = .480, SE = .196, *t* = 2.45, *p* = .014). [↑](#footnote-ref-7)
7. One reason for the lack of a quadratic effect in this instance may be the context examined. TED speakers are generally skilled communicators and thus may not add a great deal of superfluous hand movement. [↑](#footnote-ref-8)
8. Note, merely holding an object doesn’t automatically mean a gesture is a highlighter. The nature of the movement must always be considered in relation to the speaker’s words. Raising a bottle while saying “this is a water bottle,” for example, directs attention to the object, classifying the gesture as a highlighter. In contrast, using the bottle to mimic the action of pouring water into a glass while saying “so I filled the glass” would be an illustrator. [↑](#footnote-ref-9)
9. If someone scratched an itch while saying that they wondered if they walked through poison ivy, though, that movement would be related to what they were saying (and would be a highlighter). [↑](#footnote-ref-10)
10. Note that unrelated movements differ from contradictory ones. If a speaker points at a phone while saying “see this bottle,” the gesture would not only be unrelated to the spoken content, but it’d also directly contradict it. Such contradictory movements are unusual, though, and were not observed in the field data, so are not discussed further. [↑](#footnote-ref-11)
11. For example, Asalıoğlu and Göksun (2023) found no consistent evidence that illustrators signal greater emotional intensity from the speaker compared to highlighters or speech only. [↑](#footnote-ref-12)
12. Note that not all illustrators may equally signal competence. Metaphorical illustrators may be especially effective as they require greater communicative fluency and conceptual mastery. While even simple iconic gestures can aid clarity, and thus signal that the speaker is knowledgeable, metaphorical gestures may more strongly convey expertise by showing that the speaker understands the topic deeply enough to explain it in an accessible way (see the General Discussion for further elaboration). [↑](#footnote-ref-13)
13. Gemini is not able to watch videos, so each video snippet is represented as a temporal sequence of frames. [↑](#footnote-ref-14)
14. Illustrators led audiences to respond more favorably (*b* = .030, SE = .011, *t* = 2.67, *p* = .008), while highlighters (*b* = .026, SE = .025, *t* = 1.05, *p* = .296), unrelated movements (*b* = .016, SE = .022, *t* = .74, *p* = .460), and no movement (*b* = .011, SE = .021, *t* = .52, *p* = .602) did not have the same positive effect. [↑](#footnote-ref-15)
15. No Movement serves as a baseline, allowing us to isolate the effects of hand movements, regardless of their communicate intent. Unrelated movements help isolate the effects of the specific gestural communication intent. [↑](#footnote-ref-16)
16. Audio analysis using the *librosa* Python package revealed no differences on vocal-related dimensions across conditions (e.g., vocal pitch: 47.2 vs. 46.7 vs. 47.2. vs. 46.6, intonation: .19 vs. .18 vs. .19 vs. .18, loudness: .03 vs. .03 vs. .03 vs. .02, pausing: 0 vs. 0 vs. 0 vs. 0, recording duration: 11 seconds vs. 11 seconds vs. 11 seconds vs. 11 seconds). Video analysis using Face++ revealed no differences on video-related features and facial expressions (e.g., quality: 85 vs. 86 vs. 84 vs. 83, smile: .11 vs. .09 vs. .12 vs. .10, facial neutrality: 42 vs. 49 vs. 51 vs. 50, surprise: 15 vs. 12 vs. 15 vs. 18). [↑](#footnote-ref-17)
17. We mistakenly did not pre-register this measure. [↑](#footnote-ref-18)
18. Given our interest in movement type’s effect, the pre-registered analyses focused on that, and the study was not powered to expect significant effects within each individual pitch. That said, insignificant movement type x stimulus set interactions indicate that movement type has comparable effects across all three pitches (understandability F(6, 880) = 1.47, *p* = .185, competence F(6, 880) = .83, *p* = .545, and persuasion F(6, 880) = 1.90, *p* = .078). [↑](#footnote-ref-19)
19. Illustrators and complexity are not related (*r* = .01, *p* = .62), upholding the assumption of orthogonality (Cohen et al. 2003). Alternate measures of language complexity (i.e., Flesh-Kincaid) find similar results (*b* = .022, SE = .012, *t* = 1.92, *p* = .055). [↑](#footnote-ref-20)
20. These controls further ensure that the measured hand movement reflects the speaker’s actual movement and are not artifacts of camera framing or cinematographic effects. [↑](#footnote-ref-21)
21. The dependent variable engagement rate is truncated in the interval [0, 1], so we adopted censored Tobit model. We did not include log(views) as a covariate in this model. [↑](#footnote-ref-22)
22. Given that we have nearly one observation per speaker, speaker fixed effects were not included. [↑](#footnote-ref-23)
23. Some participants (N = 5) failed a pre-task asking to transcribe what a person was saying in a 4-seconds video, and almost 30% of respondents answered “no” to the question asking whether they were able to see the video. The videos format was optimized for Chrome, but many participants used other browsers, which contributed to the high failure rate. That said, the failure rate was similar across conditions (i.e., 4.8% for no movement, 5.5% for illustrators, 6.3% for unrelated, and 7.8% for highlighters). [↑](#footnote-ref-24)
24. Some participants (N = 43) mentioned technical issues in the comments section at the end of the survey (e.g., “could hear the audio but the video did not play”), while others (N = 84) failed a pre-task asking to transcribe what a person was saying in a 4-seconds video. [↑](#footnote-ref-25)
25. Confirming that the speaker kept all the rest constant across conditions, audio analysis revealed no differences on vocal-related dimensions across conditions (e.g., in the learning app pitch, vocal pitch: 46.8 vs. 47 vs. 46.6. vs. 46.9, intonation: .18 vs. .17 vs. .17 vs. .18, loudness: .02 vs. .03 vs. .02 vs. .02, pausing: 0 vs. 0 vs. 0 vs. 0, recording duration: 12 seconds vs 12 seconds vs. 12 seconds vs. 12 seconds) as well as video analysis revealed no differences on video-related features and facial expressions (e.g., in the water system pitch: quality: 87 vs. 87 vs. 84 vs. 83, smile: .07 vs. .09 vs. .13 vs. .08, facial neutrality: 52 vs. 39 vs. 61 vs. 59, surprise: 14 vs. 9 vs. 15 vs. 26). [↑](#footnote-ref-26)
26. We pre-registered separate reporting of the dependent variables and decided to measure purchase intention after preregistering the study. Principal components analysis finds that speaker liking, product liking, and purchase intention all form a single component (all loadings > .86), so we combined them into a single index as in Study 3a. [↑](#footnote-ref-27)
27. We initially pre-registered movement engagingness as a potential mechanism, but upon reflection, we realized it provides limited insight into what is actually going on. Even if differences were observed across gestures, this measure would not explain *why* certain gestures are more engaging. Furthermore, if an illustrator is more engaging than another type of gesture, it would not be surprising if this led to more favorable impressions about the speaker (e.g., how competent they seem) and other positive outcomes. [↑](#footnote-ref-28)
28. This may be driven by what this variable captures. While gesture types may shape audience’s evaluations of what they have seen (i.e., the communicator and product), their willingness to recommend the product to others may be driven more by the product itself. [↑](#footnote-ref-29)
29. Given our focus on the effect of movement, the pre-registered analyses focused on that overall main effect. Further, examining the movement type x stimulus set interaction shows that movement type has comparable effects across all three pitches on speaker liking (F(6, 880) = 1.79, *p* = .10), understandability (F(6, 880) = 1.47, *p* = .185), and competence (F(6, 880) = .83, *p* = .545). [↑](#footnote-ref-30)
30. Participants who were not able to see the video (N = 22) or failed the attention check (N = 1) asking, “The presenter was: male or female” were excluded, leaving a final sample of 77. [↑](#footnote-ref-31)
31. While vocal cues like volume or volume variation can sometimes signal confidence (van Zant and Berger 2020), they can be interpreted multiple ways (e.g., signaling stress, dominance, or emotionality depending on the context or speaker; Hildebrand et al. 2020 for a review). Consequently, we focus on linguistic signals of certainty to isolate a cue that consistently signals confidence and provides a cleaner test of the process through moderation. [↑](#footnote-ref-32)
32. The lexicon contains 3,485 words and phrases rated on the degree of certainty they imply (0 = very uncertain, 9 = very certain; see Rocklage et al. 2023). Phrases like “is definitely” and “beyond any doubt” indicate high certainty (M = 8.59 and 8.81, respectively), for example, while “unsure” and “just don’t know” indicate low certainty (M = .90 and .63, respectively). Phrases like “probably” and “it seems so” indicate moderate certainty (M = 4.03 and 5.63, respectively). Certainty was measured by participants’ most certain expression. [↑](#footnote-ref-33)
33. Illustrators and certainty language are not related (*r* = –.02, *p* = .43), upholding the assumption of orthogonality (Cohen et al. 2003). [↑](#footnote-ref-34)
34. Certainty language didn’t interact with highlighters (*b* = .002, SE = .008, *t* = .29, *p* = .774), unrelated movements (*b* = .007, SE = .008, *t* = .96, *p* = .337), or no movement (*b* = .009, SE = .007, *t* = 1.18, *p* = .681). [↑](#footnote-ref-35)