Facial Expressions Modulate the Interpretation of Spoken Quantifiers in Communication

Yichen Wu

Submitted in fulfilment of the requirements for
the Degree of Master of Science
School of Psychology and Neuroscience
University of Glasgow
August 2022

Abstract

Face-to-face communication is multimodal and involves different information channels between interacting individuals, including speech, hand/arm gestures, vocalizations, and facial expressions. However, due to methodological challenges, most studies of multimodal communication have focused primarily on (1) hand/arm gestures while neglecting facial expressions and (2) the production of these gestures rather than their perception and interpretation. Consequently, it remains unknown how facial expressions contribute to multimodal communication, including how they interact with speech. We addressed this knowledge gap by investigating whether and how facial expressions influence the interpretation of spoken utterances of vague quantifiers (e.g., *many* or *several*). In each trial, participants viewed different faces identities who each uttered a sentence using a vague quantifier—e.g., 'Of these, several are cows'— while displaying one of two facial expressions—opening or closing— or neutral. We expected that the opening facial expression would lead to *larger* number responses and the closing facial expression to lead to *smaller* number responses, analogous to hand gestures. We therefore examined whether participant number estimations shift in these expected directions. Results show that facial expressions modulated the participants' number responses as expected. Specifically, four out of ten participants' responses increased with opening facial expressions and decreased with closing facial expressions. This result suggests that facial expressions can represent quantities in a similar way as iconic hand gestures. Further, two participants increased their responses regardless of facial expression type, two participants increased their responses only for opening facial expression, and two participants showed no significant effect. Together, these results suggest that facial expressions serve several pragmatic functions in communication by operating as iconic gestures or as emphasizers. To our knowledge, this is the first study showing that facial expressions can influence the interpretation of vague quantifiers. Our results lay a foundation for future work examining how people interpret multimodal signals in daily conversation.

Contents

1 Summary

Face-to-face communication is multimodal. It involves different channels of information passing between signallers and receivers, including speech, hand/arm gestures, vocalizations, and facial expressions (e.g., Holler and Levinson, 2019). Due to methodological challenges, most of the relevant studies have focused primarily on (1) hand/arm gestures while neglecting facial expressions (Bavelas and Chovil, 2018) and (2) the production of signals rather than the perception and interpretation of them. Consequently, it remains unknown how facial expressions contribute to multimodal communication and influence the interpretation of information between communicators.

We aim to bridge this knowledge gap by investigating whether and how facial expressions influence the interpretation of spoken utterances. Specifically, we studied whether facial expressions influence the estimation of numbers similar to number-relevant iconic hand gestures. Some hand gestures are iconic—that is, they resemble referents (what these gestures try to represent). For example, a small gap between the thumb and index finger can represent numbers such as 'small/tiny number' as opposed to 'large/huge number' (Woodin *et al.*, 2020). Observational studies show that people produce facial expressions similar to iconic gestures, such as tightly squeezing the eyes to represent a 'small number' or raising the eyebrows to represent a 'large number' (see video examples in Woodin *et al.*, 2020). However, no study has shown whether facial expressions can be perceived similarly to iconic gestures.

We tested this empirically by asking participants to estimate the number of objects (e.g., cars) a speaker referred to when using one of three quantifiers—'*several*', '*few*' and '*many*' (e.g., "of these, *many* were grey"). The speakers uttered these quantifiers while displaying one of two facial expressions—'opening' (wide opening eyes) and 'closing' (squinting)—or a neutral face. We generated videos of the above facial expressions on speaking faces using a generative model of the human facial movements (Generative Face Grammar—GFG; Yu, Garrod and Schyns, 2012), comprised of 42 individual facial movements called Action Units (AUs; Ekman and Friesen, 1978) that can be activated and combined to create any biologically plausible facial expressions. We tested 10 white Western participants, each completing 540 trials. To examine whether Opening facial expression can increase and closing facial expression can decrease the participants' number responses compared to the neutral face, we modelled the data with linear mixed models. Specifically, we investigated whether participants' number responses to the opening/closing facial expression trials differed significantly from the neutral face trials.

Results show that, for 4/10 participants, opening and closing facial expressions modulated their number responses in the expected direction—the opening facial expression led to higher number responses and the closing facial expression led to lower number responses compared to a neutral face. We also observed individual differences. Two participants increased their number responses when viewing either the opening or closing facial expressions compared to a neutral face. Two increased their responses only when viewing the opening facial expression compared to a neutral face; two showed no significant effect according to facial movements. These results suggest that facial expressions can serve as iconic gestures to represent quantities.

To our knowledge, this is the first study to show that facial expressions can impact the perceived semantic meanings of spoken vague quantifiers. We anticipate that our findings will enable a deeper understanding of multimodal communication that includes facial expressions.

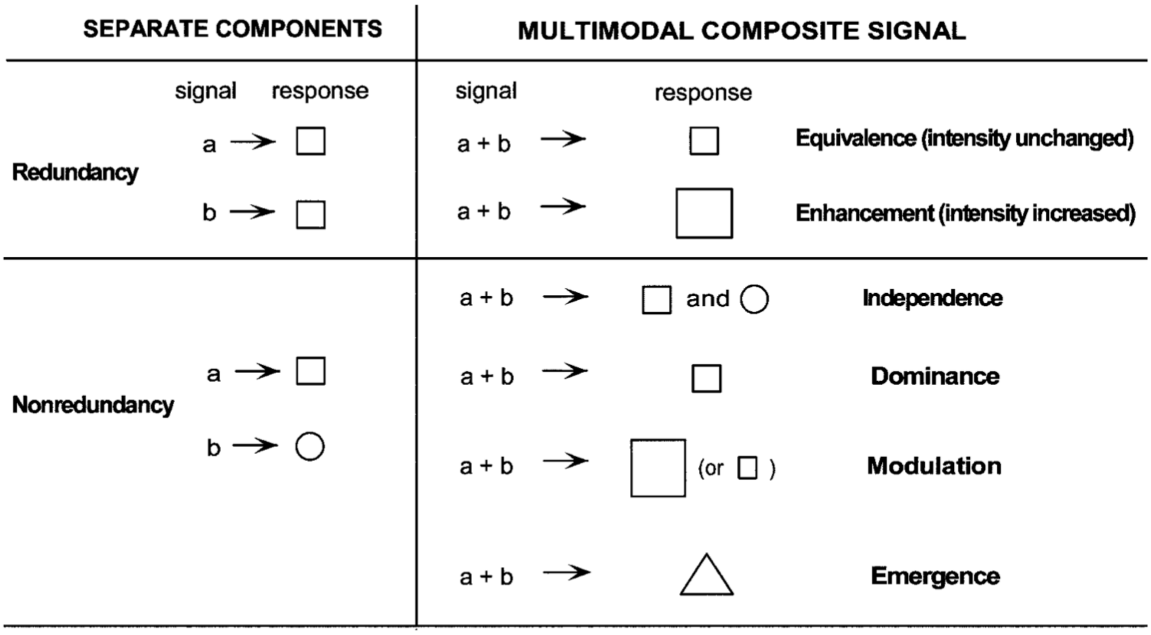
2 Introduction

More and more studies are now investigating language as a multimodal phenomenon in communication because spoken languages are embedded mostly in face-to-face conversations, which involve different channels of information going back and forth between speakers and addressees (Vigliocco, Perniss and Vinson, 2014; Perniss, 2018; Holler and Levinson, 2019). Counterintuitively, compared to unimodal signals, receiving multimodal signals in communication can facilitate the processing of information instead of hindering it. For example, when questions were asked with manual and/or head gestures, addressees answered faster than without these visual inputs—i.e., when questions were asked only as an audio message (Holler, Kendrick and Levinson, 2018). Multimodal facilitation also appeared in cases that did not involve communication. For example, the recognition of objects was faster when presented with objects' visual information and their congruent sounds, in contrast to presenting their audio or video information solely. Multimodal facilitation also exists in non-human animals. For example, jumping spiders responded to prey faster when presented with both a visual stimulus and motor stimulus—for example, vibration on the ground created by the prey— compared to each alone (Roberts, Taylor and Uetz, 2007).

In addition to speeding up the processing of information, multimodal signals can provide a variety of extra pieces of information for their receivers. There were 6 types of response change that multimodal signals can lead to compared to responses caused by the separate unichannel components (see Figure 1; Partan and Marler, 1999). For example, 'Emergence' represents situations where new responses emerge when different modal signals are combined. One of the famous instances of emergence is the McGurk effect: when the auditory component (sound of 'Ba-Ba') synchronizes with the lip movements of 'Ga-Ga', the receiver tends to perceive the word as 'Da-Da' rather than 'Ba-Ba' or 'Ga-Ga' (McGurk and MacDonald, 1976).

Figure 1

How Multimodal Signals Can be Integrated



Note. Adapted from *Communication Goes Multimodal*, by S. Partan and P. Marler, 1999, p. 1272. Copyright 1999 by American Association for the Advancement of Science.

2.1 Gestures and Facial Expressions in Communication

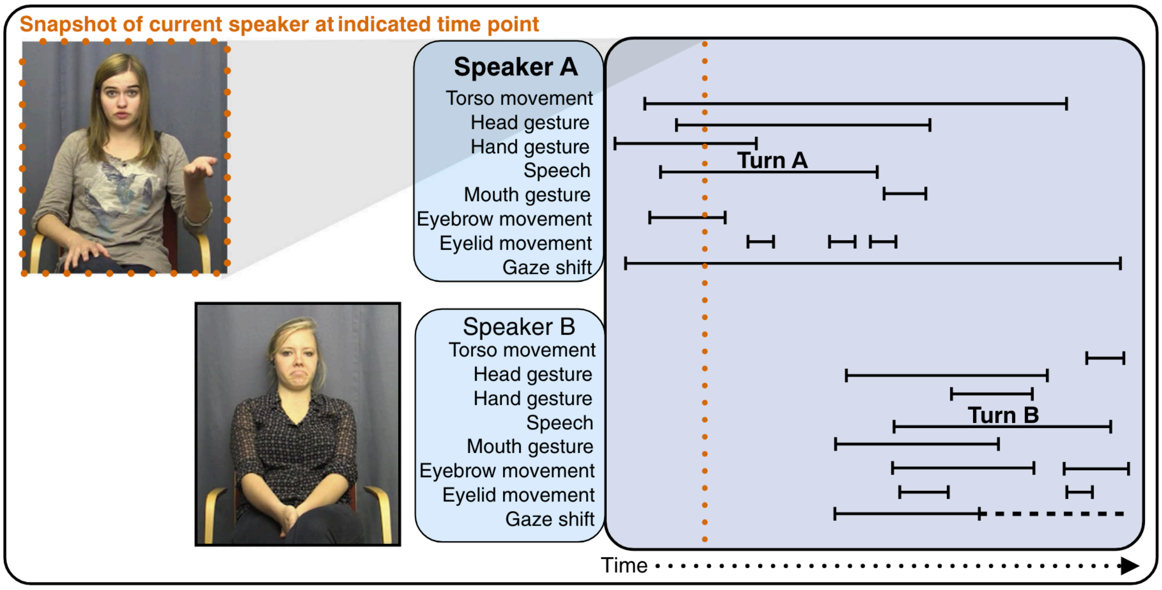
In face-to-face conversation, facial expressions and gestures are almost inevitably involved and form part of the complex multimodal signals in a visual + vocal data stream (Perniss, 2018). For example, in turn-taking conversations (see Figure 2; from Holler and Levinson, 2019), utterances from two interlocutors are consistently accompanied by gestures and facial expressions, indicating the relevance of gestures and facial expressions in multimodal communication.

2.1.1 Multimodal Communication Studies

Gestures afford a wide range of semiotic signalling strategies. Ortega and Özyürek developed a coding scheme for silent gestures by investigating how gestures can communicate (Ortega and Özyürek, 2020). The coding scheme contains systematic rules/strategies derived from 109 concepts represented by gestures across all participants. As Figure 3 shows, from left to right, each gesture belongs to the 'action', 'representing', 'drawing' and 'personification' strategy, respectively. People use these four strategies to represent different concepts. For example, 'personification' is often used to represent animate entities such as 'bird'. These gestures are all iconic, meaning that they mimic the sensorimotor features of the represented concepts (Ortega and Özyürek, 2020).

Figure 2

Data Stream of Multimodal Signals in Turn-talking Communication



Note. Adapted from *Multimodal Language Processing in Human Communication*, by J. Holler and S. C. Levison, 2019, p. 642. Copyright 2019 by Elsevier Ltd

Figure 3

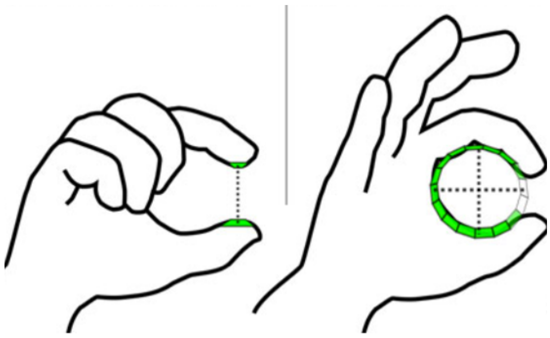
Examples of Iconic Gestures

Note. Adapted from *Systematic Mappings Between Semantic Categories and Types of Iconic Representations in the Manual Modality: A Normed Database of Silent Gesture*, by G. Ortega and A. Özyürek, 2019, p. 4. Copyright 2019 by The Psychonomic Society, Inc.

Similarly, with a more refined definition of gestural iconicity, some studies show that similar/same hand gestures can signify different gesture forms (e.g., Hassemer and Winter, 2018). Hassemer and Winter argued that gestural iconicity is not the direct mapping between the shape of the gesture and its referent—i.e., what the gesture tried to mimic. Instead, it is the relationship between the mental representation—gesture form representing some imagined spatial features such as lines and surfaces—and its referent. Specifically, they argued that the mental representations/gesture form is derived from the actual configuration of the gesture—i.e., the physical form such as hand shape and movement. Based on this idea, Hassemer and Winter then showed through their empirical data that even though the physical forms of the iconic gestures are the same, the gesture forms of the iconic gestures can be modulated by the non-iconic element. As Figure 4 shows, although the iconic parts of both gestures (the round shape made by the thumb and index finger) were the same/similar to each other, participants read them differently. When the other three fingers were curled in, they viewed it as a height gesture. In this case, the gesture form in this case is the distance between the tip of the two fingers. On the other hand, when the three fingers were not curled in, they viewed it more as a shape gesture. The gesture form, in this case, is the outline between the two fingers.

Figure 4

Two Examples of Hand Iconic Gestures



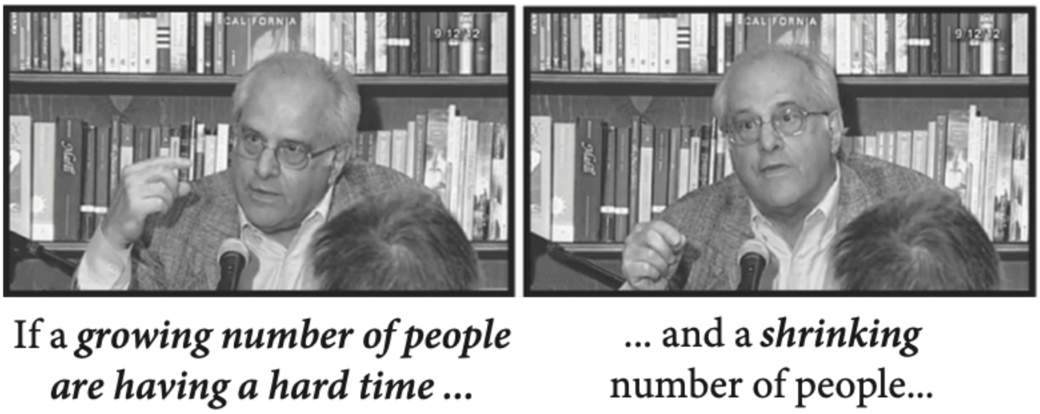
Note. The left-side gesture is the height gesture and the right-side one is the shape gesture. Adapted from *Decoding Gestural Iconicity*, by J. Hassemer and B. Winter, 2018, p. 3036. Copyright 2018 by Cognitive Science Society, Inc.

2.1.2 More Specific: Number Iconic Gestures

Certain iconic hand gestures can represent the number magnitude beyond height and shape. The reason is that people understand abstract concepts such as number by mapping them onto a more concrete domain such as physical space (Winter, Marghetis and Matlock, 2015; originated from Conceptual Metaphor Theory by Lakoff and Johnson, 1980). For example, when expressing numbers in words, people will use expressions such as ‘a tiny/huge number’ or ‘the price is rising/falling.’ Similarly, when expressing numbers using hand gestures, people will use those representing certain physical space to indicate number quantity. For example, the study by Winter, Perlman and Matlock investigated what gestures were involved when people were talking about numbers (Winter, Perlman and Matlock, 2013). They searched in TV news archives for video clips where the people mentioned ‘number’ with spatial adjectives such as ‘large’ and ‘tiny’ or when they made expressions implying the change of quantity, such as ‘a shrinking number’. After selecting 552 video clips, they adopted a qualitative approach to summarise the gestures involved. They concluded that the ‘height’ gesture mentioned in Figure 4 could indicate ‘a tiny number’. Additionally, when mentioning growing or shrinking numbers, some of the speakers also used the ‘height’ gesture. When speaking of ‘growing number,’ they moved their index finger and thumb away. On the other hand, when speaking of ‘a shrinking number’ they moved the two fingers toward each other (as shown in Figure 5).

Figure 5

An Example of Using Number-relevant Hand Iconic Gestures in Speech

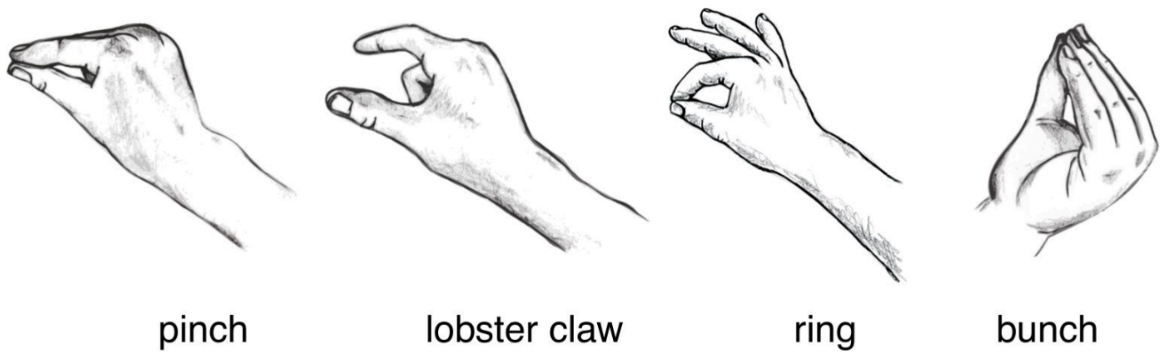


Note. Adapted from *Using Space to Talk and Gesture About Numbers: Evidence from the TV News Archive*, by B. Winter, M. Perlman and T. Matlock, 2013, p. 387. Copyright 2013 by John Benjamins Publishing Company.

Quantitative approaches have also been used to study gestures (e.g., Woodin *et al.*, 2020). Similar to the previous study (Winter, Perlman and Matlock, 2013), their data were video clips from TV News archive. However, their video clips were more specified: they all contained phrases ‘tiny/small/large/huge number(s)’. Instead of just qualitative summarising/categorizing gestures, they first broke down hand gestures into 8 parameters (e.g., Hand Configuration: Closed vs Open hand shape; see examples of closing hand in Figure 6). Then, they investigated how the adjectives (either ‘tiny’, ‘small’, ‘large’ or 'huge') in the utterance correlate with the parameters. In line with earlier work (Winter, Perlman and Matlock, 2013), results showed that when mentioning 'tiny number', the proportion of 'closing hand' hand configuration was significantly higher than when mentioning 'large number' and 'huge number'.

Figure 6

Examples of closing Hand Gestures



Note. Adapted from *'Tiny numbers' are actually tiny: Evidence from gestures in the TV News Archive*, by G. Woodin et al., 2020, p. 6. Copyright 2020 by Woodin et al.

2.2 The Current Study

Studies on iconic gestures have increased remarkably in recent years. However, there are still few studies addressing how facial expressions play a role in conversation (e.g., see Bavelas, Gerwing and Healing, 2014). Consequently, it remains unknown how facial expressions contribute to multimodal communication, including how they interact with utterances. This neglect in the literature stems mainly from two reasons. First, historically, facial expression studies have focused mostly on emotion and affect based on Darwin's theory of the evolution of facial expressions of emotion (Darwin and Prodger, 1998) which was further developed by Ekman who established the influential theory of six basic emotions (Ekman, 1971). Second, facial expressions are complex dynamic signals, which makes empirical investigation challenging. According to the Facial Action Coding System (FACS; Ekman and Friesen, 1978), the human face comprises more than 40 individual facial movements called 'Action Units' (—'AUs') that can be combined in different ways, thus resulting in thousands of possible facial expressions. In the field of empirical research, powerful tools have been developed to navigate this complexity and study facial expressions following the FACS in quantitative ways and adopted data-driven methodologies. For example, a platform based on FACS that can generate facial expressions by activating and combining AUs (see Figure 7; Yu, Garrod and Schyns,

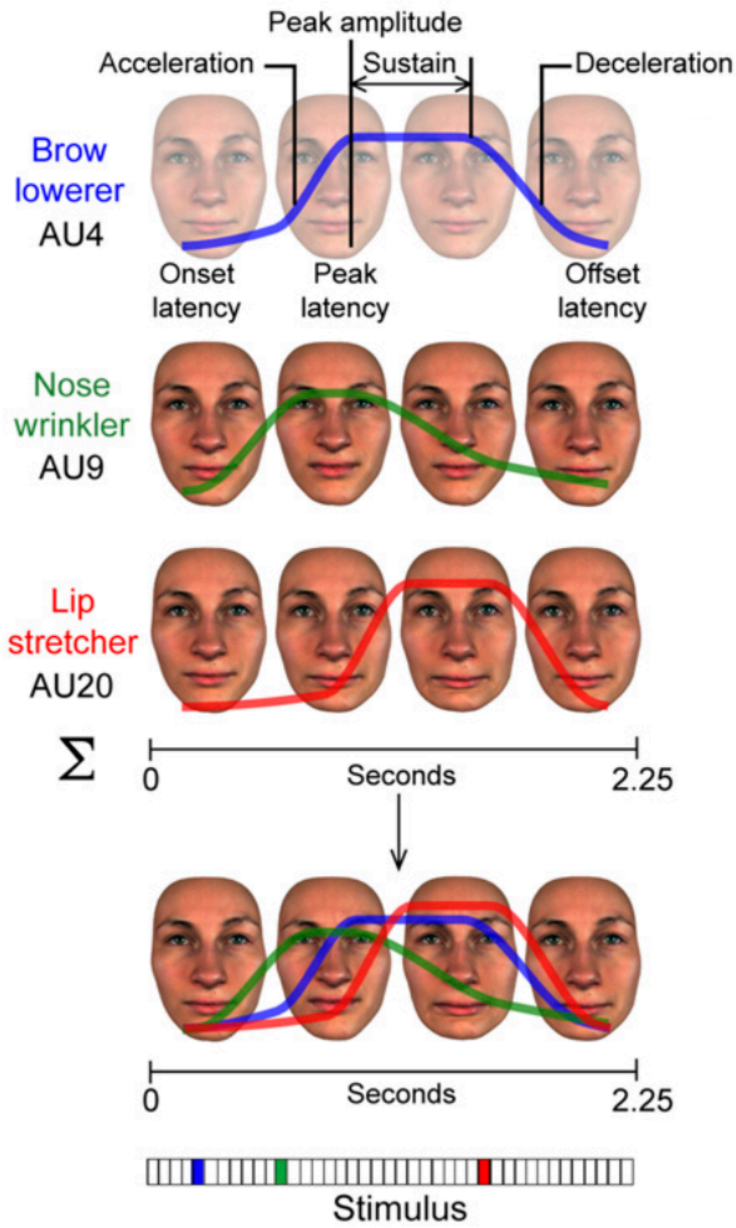
2012) with psychophysical methodologies to objectively measure the relationship between specific facial movements and participants' perception. Specifically, reverse correlation is an agnostic approach where few assumptions are made about which stimulus features will elicit a given response (Ahumada and Lovell, 1971). In contrast, classic behavioural studies use a set of pre-defined theory-driven stimuli and examine how participants respond to these stimuli. In a reverse correlation methodology, however, studies use noisy stimuli (in this case, randomly generated facial expressions comprised of combinations of AUs). The aim is to objectively examine what stimulus features participants use to resolve the perceptual task, such as categorizing emotions (e.g., Jack *et al.*, 2012), affective states (e.g., Chen *et al.*, 2018) and social traits (Gill et al., 2014; Hensel et al., 2021). Instead of investigating the theory of stimulus features, reverse correlation studies use data in such perceptual tasks to derive stimulus features. Reverse correlation can thus be used to objectively identify the stimulus features that drive perception—e.g., what specific facial movements contribute to the perception of sadness—without being contaminated by assumptions made about which stimulus features will drive response (see Jack and Schyns, 2017).

On the other hand, studies of facial expressions in the language sciences have mostly performed conversational analyses that primarily comprise qualitative and observational approaches—for example, counting the occurrences of certain facial expressions in certain contexts (Bavelas and Chovil, 2018). These studies have mostly focused on the production of facial expressions but not the perception of them and hence do not address the full picture of conversations between signallers and receivers. That being said, the current literature highlights the importance of facial expressions in face-to-face multimodal communication and their impact on comprehension. One of the first influential multimodal studies showed that synchronized lip movements ('Ga-Ga') changed the perception of the sound (from 'Ba-Ba' to 'Da-Da'; McGurk and MacDonald, 1976). Further, upper facial movements, such as frowning with raised inner eyebrows, support speech acts of questions and are considered more coherent with questions (by participants), while frowning with wide-opened eyes supports speech acts of requests/orders and is considered more coherent with requests/orders (Domaneschi et al., 2017). Together, this suggests that facial signals serve as co-speech gestures to impact comprehension, and highlights the importance of examining communication as a multimodal phenomenon to provide a unified and holistic account (Holler and Levinson, 2019).

In this study, we aim to bridge the knowledge gap in multimodal facial expression research in communication by using a perception-based approach that can objectively and precisely examine whether and how certain facial expressions influence the interpretation of spoken utterances. Specifically, and as discussed in the previous section, co-speech gestures can iconically depict and support spoken language. We aim to investigate whether facial expressions can serve a similar role and can be understood as 'facial gestures' in a way. Previous corpus data showed (without further investigating) that the utterances are accompanied by facial expressions when mentioning 'tiny/small number' or 'large/huge number' (Woodin *et al.*, 2020). As Figure 8 shows, some speakers raised their eyebrows when mentioning 'large/huge number' (see left column) and squinted their eyes when mentioning 'small/tiny number' (see right column). What makes opening and closing facial movements iconic? According to Darwin's theory of the evolutionary origins of facial expressions (Darwin & Prodger, 1998), facial expressions evolved from behaviours that benefit the expressor to modulate sensory exposure to social signals, including ritualized, i.e., exaggerated displays. Following this theory, Susskind and colleagues showed that fear and disgust facial expressions—opening and closing facial movements, respectively—can modulate sensory exposure (Susskind et al., 2008). Similarly, squinting vs widening the eyes is often performed when looking at small vs large objects to enhance visual sensitivity and acuity, respectively (Lee et al., 2014). Therefore, the iconic facial expressions might originate from the fact that the closing facial expression benefits the perception of small objects, and the opening facial expression benefits the perception of large objects.

Figure 7

Facial Expression Synthesis Based on Brow Lowerer, Nose Wrinkler and Lip Stretcher



Note. An example of stimulus generation. Adapted from *Distinct facial expressions represent pain and pleasure across cultures*, by C. Chen et al., 2018, p. E10014. Copyright 2018 by C. Chen et al.

Based on the facial movements we qualitatively observed in the naturalistic video corpus, we created three conditions: no facial expressions (no facial movements except mouth movements synchronized with the utterance) vs opening facial expression (Inner Brow Raiser, Outer Brow Raiser and Upper Lid Raiser) vs 'closing' facial expressions (Brow Lowerer, Cheek Raiser and Lid Tightener). To test whether such facial expressions serve the pragmatic function of iconic representation, we examined how perceivers interpret spoken utterances of vague quantifiers while speakers display these facial expressions.

Using a behavioural experiment, participants saw a series of speakers uttering vague quantifiers such as *several*, *few* and *many* in several contexts while they displayed utterance-synchronized facial expressions comprising opening and closing movements, and they estimated the numbers the speaker referred to by selecting a specific number.

We examined whether and how facial expressions impact the estimation of vague quantifiers because existing work shows that such estimations can vary across contexts. That is, vague quantifiers are inherently imprecise—for example, the quantifier ‘sometimes’ can represent frequencies ranging from 1–3 times per week (Kennamer, 1992)—and are thus subject to context variables. For example, the proportion inferred from a vague quantifier (‘many’, ‘several’, ‘few’) can vary by set size, i.e., the total number of objects being described (Newstead, Pollard and Riezebos, 1987). Cross-modal studies have also shown that visual information can influence the choice of vague quantifiers. For example, regardless of the number of the target object, the vague quantifiers ‘few’ and ‘a few’ are preferred more than others (e.g., ‘many’, ‘lots of’) when the total number of objects (targets + distractors) is low and vice versa for ‘many’ and ‘lots of’ when the total number of high. Similarly, ‘few’ and ‘a few’ are considered more appropriate than others (e.g., ‘many’, ‘lots of’) with higher numbers of non-target objects and vice versa for ‘many’ and ‘lots of’ (Coventry et al., 2010). Therefore, given that the interpretation of vague quantifiers is subject to contextual factors and that certain facial expressions are often displayed when speakers use them, we investigated whether vague quantifiers are also modulated by facial expressions.

Based on the observational data in previous studies, we hypothesized that the opening facial expression would shift participant quantity estimates upwards. In contrast, the closing facial expression would shift participant quantity estimates downwards, thus decreasing the number perceived from the same quantifier. Another possible explanation is that such opening and closing facial expressions could act as an emphasizer, in line with gesture studies. For example, precision grip gestures, such as the ring gesture shown in Figure 6, are used to represent the quantity/number/size of the objects being described and can emphasize the specific words the speaker wants to highlight (Winter et al., 2013). By extension, facial expressions could serve as an emphasizer in a similar way to highlight and stress specific words. In this case, both the opening and closing facial expressions would have a similar effect on number estimations by emphasizing the quantifier and shifting judgments up or down accordingly.

Figure 8

Examples of Potentially Iconic Facial Expressions While Speaking

"...a **large** number..."

"...a very **small** number..."

"...a **huge** number..."

"...a **tiny** number..."

Note. All video clips in Woodin *et al.* (2020) available at https://osf.io/dncjg/

3 Methods

All documents relevant to this project are available at https://osf.io/bk45f/.

3.1 Participants

We recruit 10 white British (mean age = 24.30, SD = 5.76; 5 females and 1 non-binary) to control the potential ethnicity effect. We recruited our participants via either Subject Pool (a webpage set up for participant recruitment; https://participants.psy.gla.ac.uk) or social media (e.g., Facebook). We collected participants' demographic information in Subject Pool, except for one participant recruited from social media. We recruited participants 18-35 years old, without learning difficulties or any psychiatric, psychological or neurological condition that could affect the processing of visual information (e.g., synaesthesia) or hearing (e.g., cochlear implant). We obtained participants' written informed consent before the start of the study and paid each of them £6/h. The experiment was approved by the University of Glasgow, College of Science and Engineering Ethics Committee (Application number: 300160203).

3.2 Stimuli

3.2.1 The Multimodal Stimuli of This Study

As we investigated facial expressions in different communication contexts, we combined the facial expressions with different synchronized recordings (see an example stimulus at: https://osf.io/bzt7k). We recorded two native British English speakers, one for each sex. They read the following text 'There are 60/108 animals in the field' and 'Of these, *few*/*several*/*many* are cows/chickens/horses.' These variations resulted in 2 Set Sizes × 3 Quantifiers × 3 Animal Variants = 18 contexts in total. We chose the two set sizes—60 and 108—from Newstead, Pollard and Riezebos's (1987) study because they fit in the context of '…animal in the field'. We chose the three quantifiers because, compared to other quantifiers, they cover a wider number proportion range (see Newstead, Pollard and Riezebos, 1987). We chose three common animals that fit the context of '60/108 animals in the field'. To avoid distraction in our experiments, we only presented the first recording—'There are 60/108 animals in the field'—in a text format so that participants could be more focused on the following multimodal stimulus where we manipulated the face. The second recording—'Of these, *few*/*several*/*many* are cows/chickens/horses.'—was paired with faces with synchronized mouth movement. We performed the generation of faces with utterances using Generative Face Grammar (the platform to generate facial expressions; Yu, Garrod and Schyns, 2012) and VOCA (a python package to generate lip movements of utterances; Cudeiro *et al.*, 2019). We coded facial movements based on Facial Action Coding System (FACS; Ekman and Friesen, 1978), which defines basic facial movements called Action Units (AUs). The duration—i.e., the length of the utterance/recording—of each stimulus was around 3-4s, while the dynamic facial expression durations were always 1.25s. To avoid interference between facial expressions and mouth movements, we only investigated eye-related AUs. Specifically, we defined faces with AU1-2 (Inner Brow Raiser and Outer Brow Raiser) and AU5 (Upper Lid Raiser) as the opening facial expression; and faces with AU4 (Brow Lowerer), AU6 (Cheek Raiser) and AU7 (Lid Tightener) as the closing facial expression. Such facial expressions have been implicated in displays of emotion — for example, opening (fear, surprise, high arousal) and closing (disgust, anger, sad and rejection, negative valence; Ekman, 1993; Matsumoto et al., 2008; Susskind et al., 2008; Jack et al., 2016).

Additionally, they have also been implicated in displays of social traits — for example, Brow Raiser (submissiveness, incompetence and trustworthiness) and Brow Lowerer (cold, competence, dominance and untrustworthiness; Gill et al., 2014; Hensel, 2022). As we aimed to investigate how the two types of facial expressions can modulate the perceptions of the quantifiers, we first centred the peak of the AUs right at the start of the quantifier while allowing a certain level of variation to make the faces appear naturalistic. Second, we had another set of 'neutral' faces as a control condition where there were only mouth movements with the utterance.

The stimuli generation was participant-wise. To keep the number of faces balanced across set sizes, quantifiers and animal variants for each participant, we had 540 trials. We calculated the number of trials from the following equation: [Neutral trials: 3 Variants (cows/chickens/horses) × 3 Quantifiers (*few*/*several*/*many*) × 2 Set Sizes (60/108) × 1 Expression (neutral) × 2 Genders (female/male) × 3 Repetitions = <u>108 trials</u>; Expression trials: 3 Variants × 3 Quantifiers × 2 Set Sizes × 2 Expressions (opening/closing facial expression) × 2 Genders × 6 Reps = <u>432 trials</u>; In total: 108 + 432 = <u>540 trials</u>]. The 540 trials were divided into 2 sessions, each containing 3 blocks of 90 trials. The trials were blocked by set sizes and quantifiers, resulting in 6 blocks of 90 trials. We shuffled all other variables mentioned in the above equation within each block. Additionally, for each trial, AU amplitude was randomly sampled from [0.2, 0.4, 0.6, 0.8, 1.0] with 1.0 representing the maximum amplitude (so that participants could not easily guess we were manipulating the amplitude and expect a certain effect of the amplitude). The identities also varied across trials. First, the facial identity of every trial was different. Second, ten voice identities were generated and randomly shuffled within each gender.

To familiarise participants with the study, we also prepared 6 practice trials before the start of the first block of the first session. The practice trials were the same across all participants and had a different scenario than formal trials: '60/108 cars were produced that day'; 'Of these, *few*/*several*/*many* were black/white/grey'.

3.2.2 Short Introduction to GFG

To generate the stimuli, we used a generative model of human facial movements called the Generative Face Grammar (GFG; Yu, Garrod and Schyns, 2012). In this way, we are more precise about what facial movements we should present and can vary facial identities freely without overlapping with real-life facial identities. In GFG, the default length of the facial expressions is 1.25s (30 frames per second). The generation of facial expressions is based on Facial Action Coding System (FACS; Ekman and Friesen, 1978). FACS defines basic facial movements called Action Units (AUs) from an anatomical perspective. Specifically, each AU involves the activation of unique facial muscles —for example, AU1, Inner Brow Raiser, is controlled by the frontalis and pars medialis facial muscles. The GFG can present whatever combinations between these AUs. Instead of just on/off— i.e., whether a certain AU will be presented or not, every AU had 8 parameters. Each parameter, from different aspects, controls how facial expressions are presented. For example, Peak Amp controls how strong/obvious the AU is, and Peak Latency controls when the AU reaches its maximum movement (Yu, Garrod and Schyns, 2012). Hence, using GFG, we can present almost every possible facial expression. We can also present facial expressions on different facial identities. For white ethnicity, the generative model of human facial identities we used can generate random novel facial identities based on ~400 real human facial faces (see Zhan *et al.*, 2019).

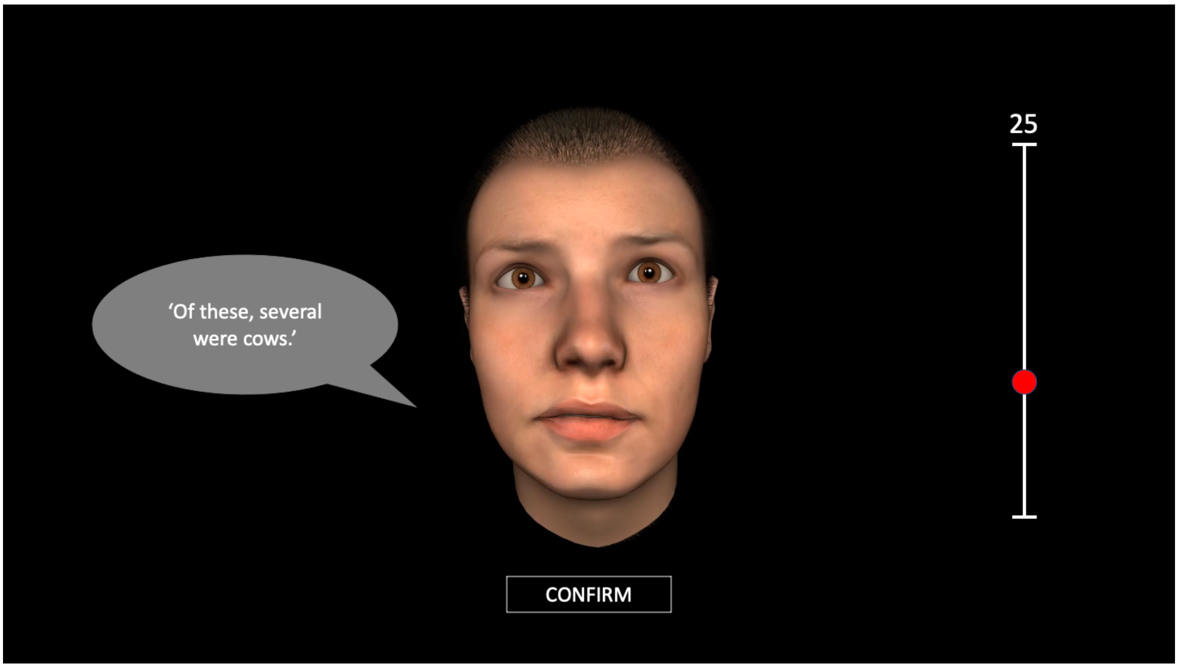
3.3 Procedure

Before the experiment, participants needed to fill in a covid form and an informed consent form. After informed consent, we instructed participants to estimate the number that the speaker on each trial was referring to. They heard the audio stimuli through a pair of headphones we provided and viewed visual stimuli on a black background displayed on a 19-inch flat panel Dell monitor with a refresh rate of 60 Hz and resolution of 1,024 × 1,280. Visual stimuli appeared in the central visual field and disappeared after it finished. We set the chin rest and the monitor to a distance of 71cm to ensure the visual angles of 14.25° vertically and 10.08° horizontally, reflecting the average face size (Ibrahimagić-Šeper *et al.*, 2006) in human social interaction (Edward, 1966). We used a chin rest to control the visual angle of the face stimulus across the experiment. This is a necessary experimental control because varying the visual angle (i.e., the size of the stimulus on the participant's retina) will alter the stimulus information that is available to the participant's visual system and thus what the participant can or cannot use to perform the task (e.g., see Smith & Schyns, 2009).

They started with 6 practice trials, during which they changed the headphone volume to the extent they saw fit. After the practice trials, they started formal trials. We blocked trials by quantifiers and set sizes and randomized trials within each block for each participant. Before each trial, there was a sentence (e.g., 'There are 60 animals in the field') presented for at least 1 second in the centre of the screen as a prompt. Participants moved on to the trial by pressing the 'PROCEED' button. After pressing the button, a fixation cross stayed for 0.5 s, followed by the facial-utterance stimulus. The audio stimuli stopped, and the visual stimuli disappeared after they finished. Then, participants made their responses. They needed to indicate the number of objects/animals using a vertical presented slider on the right of the screen and to click the 'CONFIRM' to move on to the prompt of the next trial (Figure 9). We controlled the stimulus presentation using Python 3.6.12. Participants had a 4-minute break in between every block and at least a 15-minute rest between the two sessions to eliminate fatigue of participants. After finishing 540 trials, they were given payment and debriefing before leaving.

Figure 9

An Exemplar of a Trial



Note. In the middle was the talking avatar. Below was the confirm button. On the right was the response slider. On the top of the slider was the response number. The grey callout was not shown in a real trial.

3.4 Planned Data Analyses

We performed data analyses using R 4.1.1 (R Core Team, 2021), in the integrated development environment RStudio, version 1.4.1717 (RStudio Team, 2020). We used the lmerTest package for modelling, version 3.1-3 (Kuznetsova, Brockhoff and Christensen, 2017). The other packages we used were: ggplot2, version 3.3.6 (Wickham, 2016); tidyverse, version 1.3.1 (Wickham et al., 2019). We first performed modelling based on between-participant data. After that, we performed within-participant analyses for each individual participant.

We aimed to examine whether facial expressions influence participants' number estimates. Because the response data from different set sizes were of different scales, we transformed the response data into proportion data by dividing them by their corresponding set sizes before analysis. To examine whether the opening and closing facial expressions can modulate the participants' number responses compared to a neutral face, we planned to perform a linear mixed model for all- and each-participant data respectively. The within-participant model's formula was the following:

Resp_prop ~ Expression_type + Set size + Quantifier + (1|Variant)

We treated Resp_prop as a continuous variable. We treated Expression_type (closing coded as 0, neutral as 1 and opening as 2), Set size (60 coded as 0 and 108 as 1), Quantifier (*several* coded as 0, *few* as 1 and *many* as 2) and Variant (cows coded as 0, horses as 1 and chickens as 2) as categorical factors. In the linear mixed model, we treated levels coded as 0 as the default baseline. We used the vague quantifier 'Several' as the baseline because it represents higher quantities than 'few' and lower quantities than 'many' (Newstead et al., 1987)—i.e., the midpoint. For example, if the proportion for baseline 'several' was lower than that for 'many' and higher than that for 'few' (which occurred in all but one case), then the statistical relationship between 'many' and 'few' could also be indicated. We made a baseline adjustment in the model because the original number coding was somewhat arbitrary. Specifically, we set Expression_type (neutral) as the baseline to examine whether opening/closing facial expressions impact participant responses compared to no facial expression. Variant was treated as a random factor because we were not interested specifically in how Variant can have an influence (nor the differences between each level). The between-participant formula was similar to the one for per-participant, except there was one more random factor (the No. of Participant):

Resp_prop ~ Expression_type + Set size + Quantifier + (1|Variant) + (1|Participant_No)

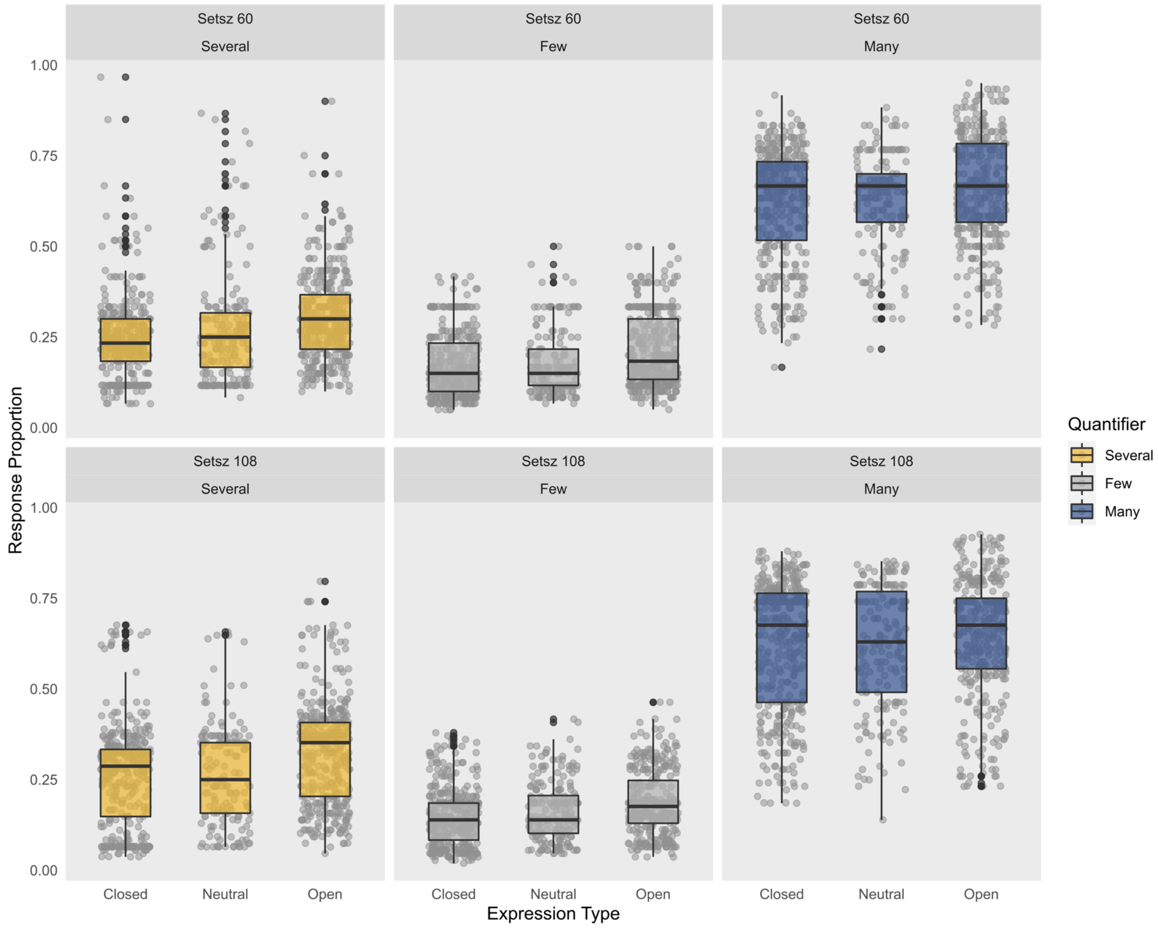
4 Results

4.1 Between-Participant Analyses

First, we ran the aforementioned between-participant model to investigate the general pattern across all participants.

Figure 10

Participants' Number Responses Under Different Facial Expressions Subset by Set Size and Quantifier for Between-participant Data



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the facial expression category: from left to right, closing, Neutral and opening facial expressions. Y-axis indicates the transformed participants' number responses: Raw responses were divided by their set size. Different panels represent different set sizes and quantifiers: rows for the former, from up to down 60 and 108; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Figure 10 shows that the effect of opening facial expression was consistent across different quantifiers and different set sizes. The effect of closing facial expression was less consistent than opening facial expression.

Table 1

Between-participant Linear Mixed Model Results

	Estimate	Std. Err	df	t value	p value
Fixed effects					
Closing	− 0.0082	0.0048	5383	− 1.73	0.084
Opening	0.036***	0.0048	5383	7.58	4.13E-14
Set size - 108	− 0.0052	0.0035	5383	− 1.19	0.14
Quantifier - *Few*	− 0.10***	0.0043	5383	− 24.97	< 2E-16
Quantifier - *Many*	0.35***	0.0043	5383	82.44	< 2E-16

	Variance	Std. Dev
Random effects		
Participant No.	0.0011	0.032
Variants	2.92E-05	0.0054

Note. * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$. Intercept and Residual were not shown in the table but included in the model.

Table 1 shows three significant results. As expected, participants on average increased their responses by 3.6% when seeing an opening facial expression, compared to a Neutral face condition, $M_{diff} = 0.036$, SE = 0.0048, $t(5383) = 7.58$, $p = 4.13e-14$, $d = 0.10$. The closing facial expression condition, though in the expected direction, was not significantly—though marginally significant—different from the neutral condition, $M_{diff} = -0.0082$, SE = 0.0048, $t(5383) = -1.73$, $p = 0.084$, $d = -0.024$. Participants' responses towards faces that spoke of '*few*' were on average 10.22% lower than compared to '*several*', $M_{diff} = -0.10$, SE = 0.0043, $t(5383) = -24.97$, $p < 2e-16$, $d = -0.34$; on the other hand, '*Many*' were on average 35.14% higher in its proportion than '*several*', $M_{diff} = 0.35$, SE = 0.0043, $t(5383) = 82.44$, $p < 2e-16$, $d = 1.12$.

To examine whether the amplitudes of the facial expressions have an influence, we conducted a new model which included the amplitude as an interaction term with the Expression_type:

Resp_prop ~ Expression_type + Expression_type:Amplitude + Set size + Quantifier + (1|Variant) + (1|Participant_No)

Table 2

Between-participant Linear Mixed Model Results (with Amplitude)

	Estimate	Std. Error	df	t value	p value
Fixed effects					
Closing	1.69E-04	0.0075	5382	0.022	0.98
Opening	0.015*	0.0075	5382	2.05	0.040
Closing:Amps	− 0.014	0.0097	5382	− 1.44	0.15
Opening:Amps	0.034***	0.0097	5383	3.57	3.62E-04
Set size - 108	− 0.0050	0.0035	5381	− 1.44	0.15
Quantifier - *Few*	− 0.10***	0.0043	5381	− 24.00	< 2E-16
Quantifier - *Many*	0.35***	0.0043	5381	82.56	< 2E-16

	Variance	Std. Dev
Random effects		
Participant No.	0.0011	0.033
Variants	2.93E-05	0.0054

Note. * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$. Intercept and Residual were not shown in the table but included in the model.

Table 2 shows the statistics of the new model with amplitude. The effect of quantifiers was relatively consistent between these two models (after including amplitudes: *few*: $M_{diff} = -0.10$, SE = 0.0043, $t(5381) = -24.00$, $p <$ 2e-16, $d = -0.33$; *many*: $M_{diff} = 0.35$, SE = 0.0043, $t(5381) = 82.56$, $p <$ 2e-16, $d = 1.12$). The effect of opening facial expression was still significant, $M_{diff} = 0.015$, SE = 0.0075, $t(5381) = 2.05$, $p = 0.040$, $d = 0.028$, while the interaction between Expression_type (opening facial expression) and AUs' amplitude was also as expected, $B_{diff} = 0.034$, SE = 0.0097, $t(5383) = 3.57$, $p =$ 3.62e-4, $d = 0.049$. Lastly, an ANOVA between Amplitude-On and Amplitude-Off models shows that adding the amplitude as a factor significantly improves the model fit ($\chi^2(2) = 14.80$, $p =$ 6.10e-4).

4.2 Within-Participant Analyses

From the previous between-participant analyses we can draw the interim conclusion that (1) in general, opening facial expression had a positive effect, which was modulated by amplitude, and (2) different quantifiers covered different proportional ranges of the number scale. However, since every participant comes with different interpretations of the messages (e.g., world knowledge/associations with the scenarios) and an idiosyncratic visual system, the influence of facial expressions could be masked by between-participant variation. Hence, in the following part, we adopted a within-participant approach. In other words, we performed our regression model for each participant respectively and combined it with quantification of the proportion of facial expression effect across all participants (Ince, Kay and Schyns, 2022).

Table 3

Significance Levels of Within-participant Linear Mixed Model Variables

	1	**2**	**3**	**4**	**5**	**6**	**8**	**9**	**10**	**11**
Closing	-	+	-	-		+		-		
Opening	+	+	+	+		+	+	+	+	
Set size - 108	-	+	+	-	+		-	-		-
Quantifier - *Few*	-	-	-	-	-	-	+	-	-	-
Quantifier - *Many*	+	+	+	+	+	+	+	+	+	+

Note. Rows show results sorted by different predictor variables, and columns represent participant No.; Colours represent the variable's significance: the brighter the colour, the lower the p-value is: '-' sign indicates a negative coefficient—i.e., decreased responses compared to the baseline. ' - ' for $p < 0.05$; ' - ' for $p < 0.01$; ' - ' for $p < 0.001$; '+' sign indicates a positive coefficient—i.e., increased responses compared to the baseline: ' + ' for $p < 0.05$; ' + ' for $p < 0.01$; ' + ' for $p < 0.001$. Intercept and Residuals were not shown in the table but included in the models.

Table 3 summarises the significance of different variables across all participants. Opening facial expression was more consistent (8 out of 10 participants increased their responses; no significant influence for 2 out of 10 participants) than its closing counterparts (4 decreased, 2 increased and 4 were not significantly influenced). Across all variables, the Quantifier '*many*' had the most consistent (larger proportion than '*several*' in 10 out of 10 participants) and the largest effect ($d_{mean} = 2.96$, SD = 2.95); The effect of quantifier '*few*' was consistent (smaller proportion than '*several*' in 9 out of 10 participants) and strong ($d_{mean} = -0.82$, SD = 0.97) as well. Notably, for Subject 8 (Participant number henceforth: Subject X), '*few*' took up a significantly larger proportion than '*several*' ($M_{diff} = 0.072$, SE = 0.0086, $t(532) = 8.34$, $p = 6.33e-16$, $d = 0.36$). The set size, on the other hand, had a more inconsistent influence across different participants—Five of the participants gave significantly lower average response proportions to set size 108 than set size 60; three gave significantly higher proportions; two gave similar proportions for both set sizes.

In the following sections, we mainly focused on the influence of facial expressions on individual participants' number responses. We examined the opening and closing facial expressions in models with and without the amplitude factor. Given the results in Table 3, we categorized the participants into 4 classes that reflect their responses: Iconic, Emphasiser, Others, and No Effect.

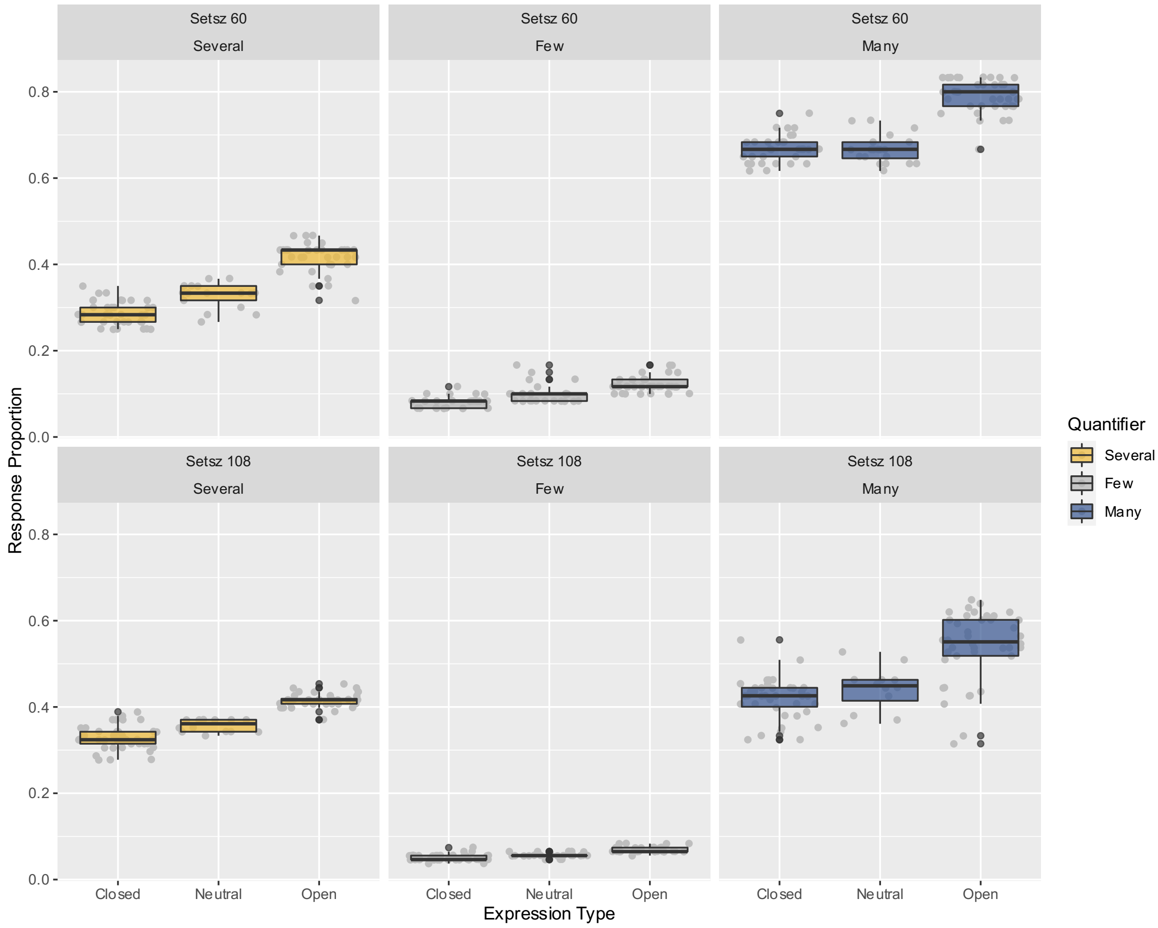
4.2.1 Responses Suggesting Iconicity

We included participants who increased their responses to the opening facial expression and decreased their responses to the closing facial expression in this class. Four out of ten participants were within this class, making it the most prevalent pattern across all participants.

Figure 11 shows that, for Subject 1, the effects of the opening and closing facial expressions were consistent across different set sizes and quantifiers. This is the case for most of the participants within this category (see Figure S1), except for Subject 9 who had the highest average response to neutral faces in the *several* × Set Size 60 panel (See Figure 12). Additionally, Subject 9 had the largest within-boxplot variance across these four participants. Subject 1 had a larger opening facial expression effect with the quantifier '*many*' compared to other quantifiers.

Figure 11

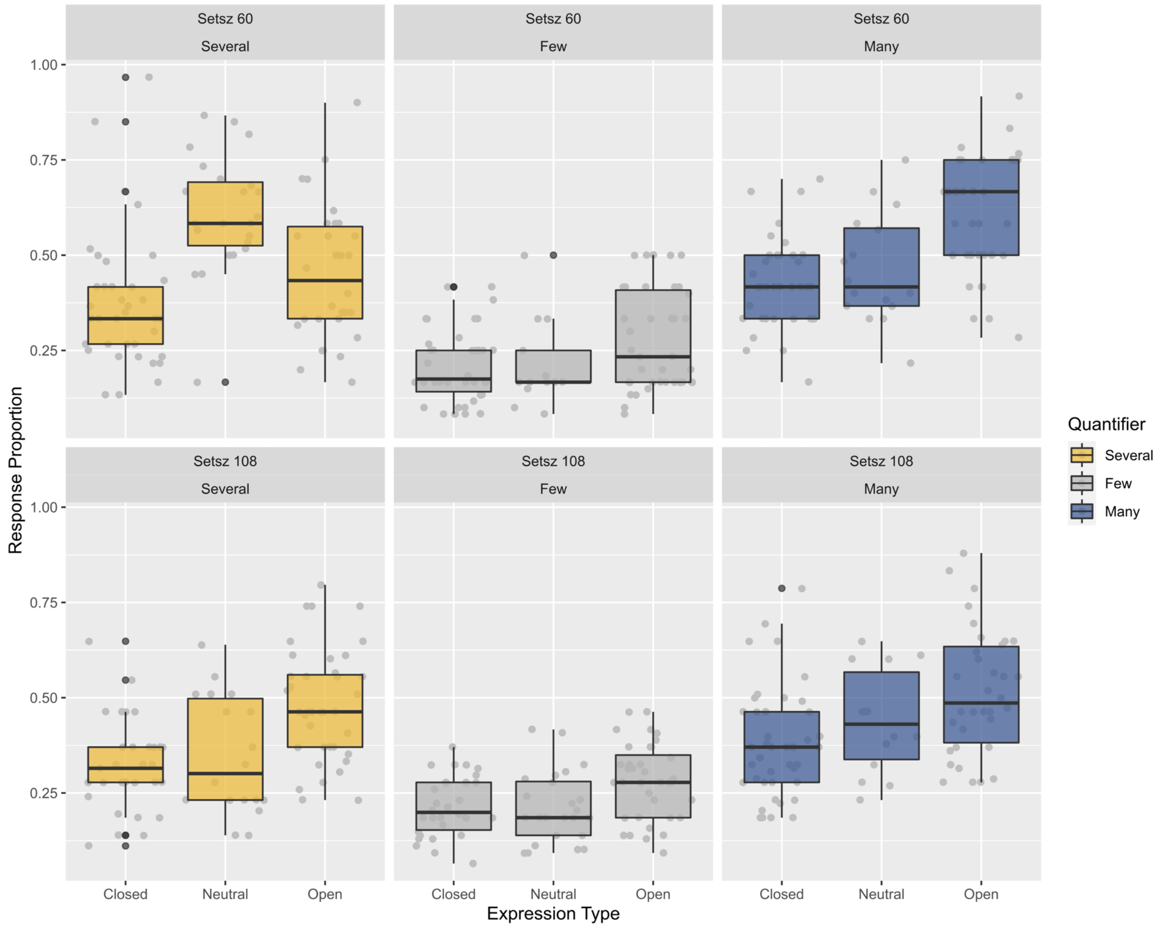
Exemplar Participant's (Subject 1) Number Responses Under Different Facial Expressions Subset by Set Size and Quantifier for Participants from Iconic Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the facial expression category: from left to right, closing, Neutral and opening facial expressions. Y axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different set sizes and quantifiers: rows for the former, from up to down 60 and 108; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Figure 12

Exemplar Participant's (Subject 9) Number Responses Under Different Facial Expressions Subset by Set Size and Quantifier for Participants from Iconic Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the facial expression category: from left to right, closing, Neutral and opening facial expressions. Y axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different set sizes and quantifiers: rows for the former, from up to down 60 and 108; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Table 4 shows that for Subject 1, opening facial expression had a stronger effect than closing facial expression (closing: $M_{diff} = -0.025$, SE = 0.0083, $t(534) = -3.033$, $p = 0.0025$, $d = -0.13$, andopeningg: $M_{diff} = 0.061$, SE = 0.0083, $t(534) = 7.38$, $p =$ 6.19e-13, $d = 0.32$); Subject 4 and 9 were the other way around: closing facial expression had a stronger effect than opening ones (Subject 4, closing: $M_{diff} = -0.033$, SE = 0.0043, $t(532.10) = -7.70$, $p =$ 6.90e-14, $d = -0.33$, andopeningg: $M_{diff} = 0.026$, SE = 0.0043, $t(532.34) = 6.05$, $p =$ 2.74e-9, $d = 0.26$; Subject 9, closing: $M_{diff} = -0.074$, SE = 0.015, $t(532.06) = -4.98$, $p =$ 8.74e-7, $d = -0.22$, andopeningg: $M_{diff} = 0.036$, SE = 0.015,

t(532.05) = 2.45, p = 0.015, d = 0.11); For Subject 3 the effect ofopeningg and closing facial expression were close (closing: $M_{diff} = -0.022$, SE = 0.0027, $t(534) = -7.99$, p = 8.63e-15, $d = -0.35$, andopeningg: $M_{diff} = 0.021$, SE = 0.0027, $t(534) = 7.73$, p = 5.22e-14, d = 0.33).

Table 4

Estimates of Closing and Opening Effects resulting from the Linear Mixed Model for all participants in the Iconic Group

	Estimate	**Std Error**	**df**	**t value**	**p value**
			Closing		
Subject 1	**− 0.025****	0.0083	534	− 3.033	0.0025
Subject 3	**− 0.022*****	0.0027	534	− 7.99	8.63E-15
Subject 4	**− 0.033*****	0.0043	532.10	− 7.70	6.90E-14
Subject 9	**− 0.074*****	0.015	532.06	− 4.98	8.74E-07
			Opening		
Subject 1	**0.061*****	0.0083	534	7.38	6.19E-13
Subject 3	**0.021*****	0.0027	534	7.73	5.22E-14
Subject 4	**0.026*****	0.0043	532.34	6.05	2.74E-09
Subject 9	**0.036***	0.015	532.05	2.45	0.015

Note. * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$.

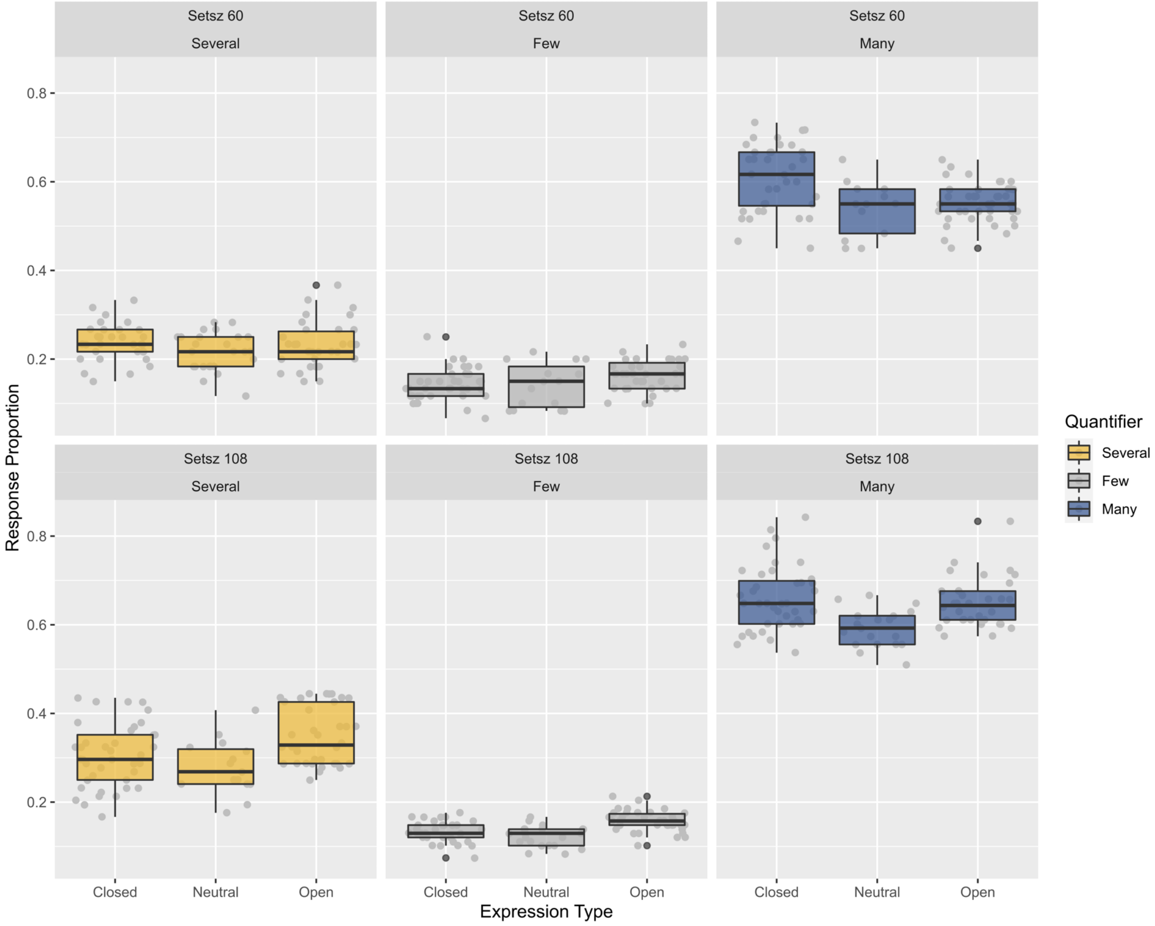
4.2.2 Responses Suggesting an Emphasiser Function

Participants in the Emphasiser group increased their responses to both the opening and closing facial expressions. Two out of ten participants fell into this category.

Figure 13 and 14 shows the two participants who increased their responses for both opening and closing facial expressions. For Subject 2, the effects were consistent across all Quantifier × Set Size panels (Figure 13). For Subject 6, on the other hand, the pattern was less consistent: for the quantifier '*few*', the effects followed the iconicity pattern (Figure 14). For '*many*', the closing facial expression had a positive and stronger effect than the opening facial expression, while for '*several*' the effects were less clear).

Figure 13

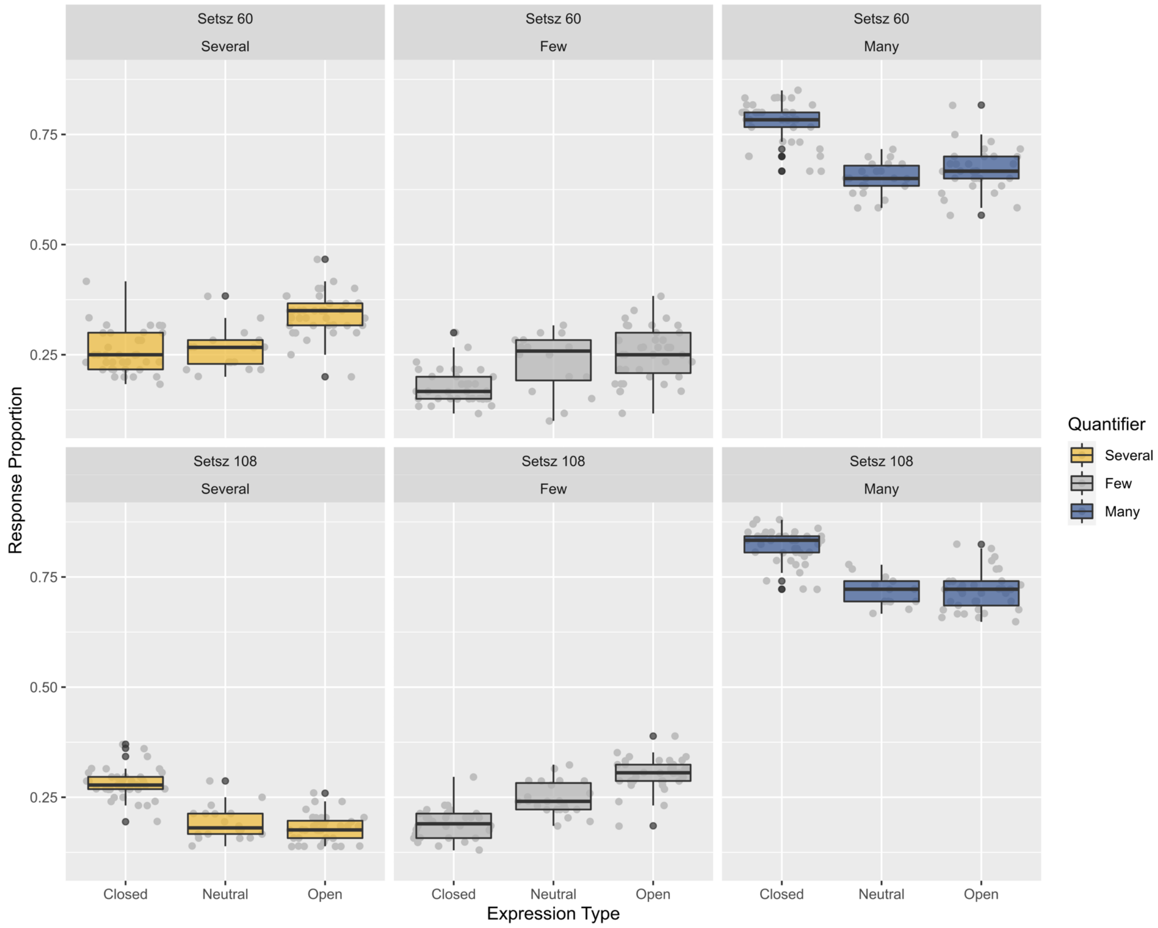
Exemplar Participant's (Subject 2) Number Responses Under Different Facial Expressions Subset by Set Size and Quantifier for Participants from Emphasiser Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the facial expression category: from left to right, closing, Neutral and opening facial expressions. Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different set sizes (upper row = 60, lower row = 108) and quantifiers (columns from left to right: '*several*', '*few*' and '*many*').

Figure 14

Exemplar Participant's (Subject 6) Number Responses Under Different Facial Expressions Subset by Set Size and Quantifier for Participants from Emphasiser Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the facial expression category: from left to right, closing, Neutral and opening facial expressions. Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different set sizes (upper row = 60, lower row = 108) and quantifiers (columns from left to right: '*several*', '*few*' and '*many*').

Table 5

Estimates of Closing and Opening Effects resulting from the Linear Mixed Model for all participants in the Emphasiser Group

	Estimate	Std Error	df	t value	p value
			closing		
Subject 2	**0.036*** **	0.0068	532.24	5.34	1.41E-07
Subject 6	**0.038*** **	0.0081	534	4.75	2.64E-06
			opening		
Subject 2	**0.035*** **	0.0068	532.16	5.24	2.29E-07
Subject 6	**0.031*** **	0.0081	534	3.82	1.50E-04

Note. * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$.

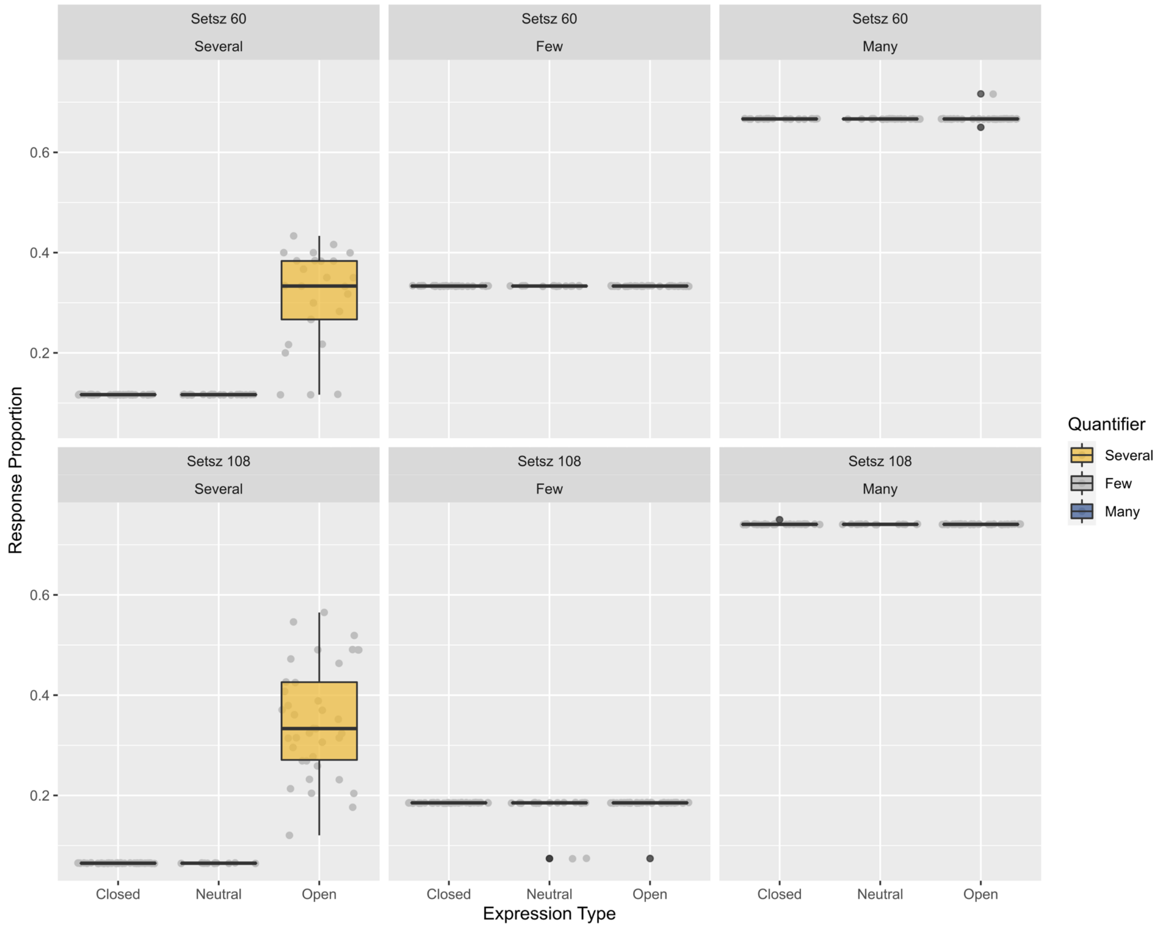
For these two participants, closing and opening facial expressions showed similar effect sizes for changing their number responses (see Table 5). For Subject 2, they increased their responses proportions by 3.6% (SE = 0.0068, $t(532.24) = 5.34$, $p = 1.41e\text{-}7$, $d = 0.23$) when seeing opening facial expressions and 3.5% (SE = 0.0068, $t(532.16) = 5.24$, $p = 2.29e\text{-}7$, $d = 0.23$) when seeing closing facial expressions. For Subject 6, they increased the proportions by 3.8% (SE = 0.0081, $t(534) = 4.75$, $p = 2.64e\text{-}6$, $d = 0.21$) and 3.1% (SE = 0.0081, $t(534) = 3.82$, $p = 1.50e\text{-}4$, $d = 0.17$) respectively to opening and closing facial expressions.

4.2.3 Opening Facial Expression Only

Figure 15 and 16 show participants who increased their responses only to opening facial expressions.

Figure 15

Exemplar Participant's (Subject 8) Number Responses Under Different Facial Expressions Subset by Set Size and Quantifier for Participants from Opening Facial Expression Group

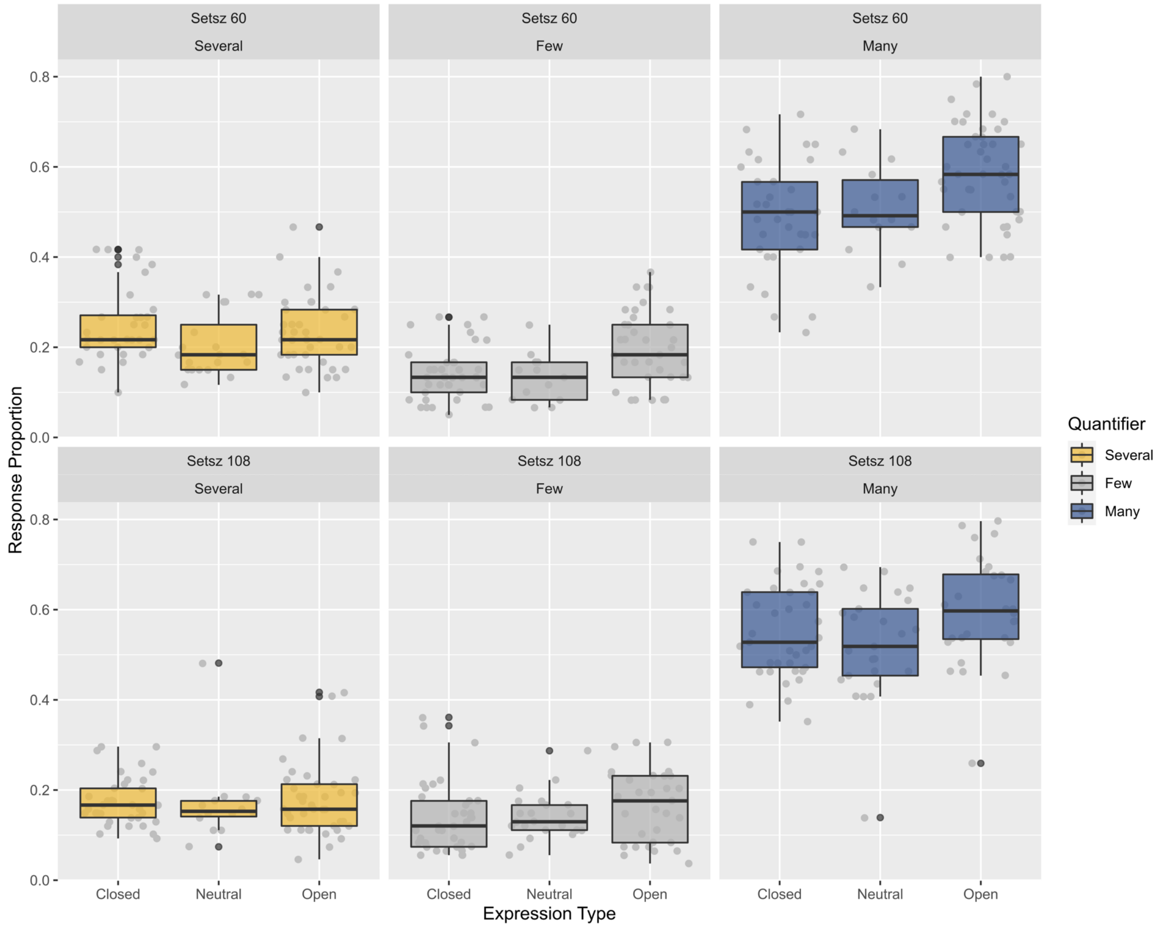


Note. Colourful boxplots and grey colour raw data. The X-axis indicates the facial expression category: from left to right, closing, Neutral and opening facial expressions. Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different set sizes (upper row = 60, lower row = 108) and quantifiers (columns from left to right: '*several*', '*few*' and '*many*').

For Subject 8, it was clear that the opening facial expression only had its effect with the quantifier '*several*' (Figure 14). At the same time, variances within other Quantifier × Set Size × Facial Expression Type boxplots were extremely low (only 1-3 values within a boxplot). For Subject 10, the positive effect of opening facial expression was consistent across different Quantifier × Set Size panels (Figure 16).

Figure 16

Exemplar Participant's (Subject 10) Number Responses Under Different Facial Expressions Subset by Set Size and Quantifier for Participants from Opening Facial Expression Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the facial expression category: from left to right, closing, Neutral and opening facial expressions. Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different set sizes (upper row = 60, lower row = 108) and quantifiers (columns from left to right: '*several*', '*few*' and '*many*').

Table 6

Estimates of Closing and Opening Effects Resulting from the Linear Mixed Model for Each Participant in the opening Only Group

	Estimate	**Std Error**	**df**	**t value**	**p value**
			Closing		
Subject 8	**5.70E-04**	0.0099	532.20	0.059	0.95
Subject 10	**0.0091**	0.011	532.12	0.85	0.40
			Opening		
Subject 8	**0.082*****	0.0097	532.10	8.52	< 2E-16
Subject 10	**0.043*****	0.011	532.22	4.04	6.22E-05

Note. * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$.

Table 6 shows that for these two participants, the closing facial expression did not influence the participants' answers significantly (Subject 8: $M_{diff} = 5.70e-4$, SE = 0.0099, $t(532.20) = 0.059$, $p = 0.95$, $d = 0.0026$; Subject 10: $M_{diff} = 0.0091$, SE = 0.011, $t(532.12) = 0.85$, $p = 0.40$, $d = 0.037$). On the other hand, participant significantly increased their responses when seeing the opening facial expression (Subject 8: $M_{diff} = 0.082$, SE = 0.0097, $t(532.10) = 8.52$, $p < 2e-16$, $d = 0.37$; Subject 10: $M_{diff} = 0.043$, SE = 0.011, $t(532.22) = 4.04$, $p = 6.22e-5$, $d = 0.18$). These two participants differed in how effective the opening facial expression was.

4.2.4 No Effect of Facial Expressions

Two participants were not influenced by either the opening or closing facial expressions (see Table 7).

Table 7

Estimates of Closing and Opening Effects Resulting from the Linear Mixed Model for Each Participant in the No Effect Group

	Estimate	Std Error	df	t value	p value
			Closing		
Subject 5	**0.0024**	0.0087	533.66	0.27	0.79
Subject 11	**0.0033**	0.0061	532	0.54	0.59
			Opening		
Subject 5	**− 0.0011**	0.0087	532.68	− 0.13	0.90
Subject 11	**0.0010**	0.0061	532	0.17	0.87

Note. * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$.

4.2.5 Amplitudes

Similar to within-participant analyses, we examined whether adding amplitudes as interaction terms with both of the facial expressions can significantly improve the model for each participant respectively. The formula we used to include the amplitude in the model for each participant was the following:

Resp_prop ~ Expression_type + Expression_type:Amplitude + Set size + Quantifier + (1|Variant)

Table 8 shows that six models showed a significantly better fit when including amplitudes. All the six models were from the eight models where either/both the closing or opening facial expressions had a significant effect. For the 6 models, including amplitudes increased the overall fit of the model but decreased the individual effect size/significance level of each facial-expression predictor. According to how the amplitude influenced the facial expression effects, the 8 models—in which either opening and/or closing facial expressions had an effect—were categorized into three groups: effects all explained by the interaction, effects partly explained by the interaction and no significant influence from the amplitude.

Table 8

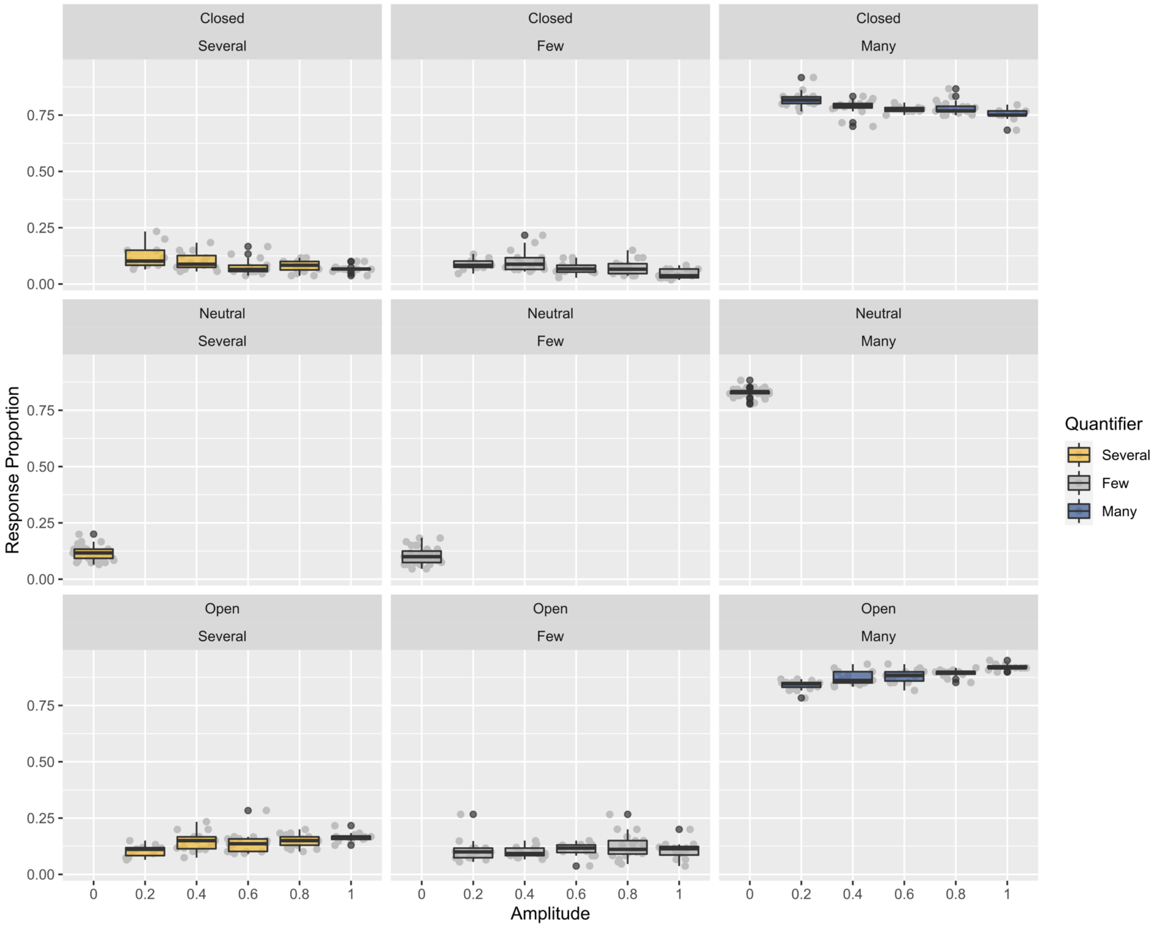
Significance Levels of Within-participant Linear Mixed Models (Both the With and Without Amplitude Ones) Variables and of ANOVA Comparisons between the Two Models

	1	**2**	**3**	**4**	**5**	**6**	**8**	**9**	**10**	**11**
Model without amp										
Closing	-	+	-	-		+		-		
Opening	+	+	+	+		+	+	+	+	
Model with amp										
Closing	-									
Opening	+		+				+			
Closing × Amp		+	-	-				-		
Opening × Amp		+	+	+		+	+	+	+	
ANOVA										
Sig amp model		***	***	***			**	***	***	

Note. Rows show results sorted by different predictor variables, and columns represent participant No. For the ‘Model’ part of the table, Colours represent the variable’s significance: the brighter the colour, the smaller the *p*-value was: ‘-’ sign indicates a negative coefficient—i.e., decreased responses compared to the baseline. ‘ - ’ for $p < 0.05$; ‘ - ’ for $p < 0.01$; ‘ - ’ for $p < 0.001$; ‘+’ sign indicates a positive coefficient—i.e., increased responses compared to the baseline: ‘ + ’ for $p < 0.05$; ‘ + ’ for $p < 0.01$; ‘ + ’ for $p < 0.001$. Intercept and Residuals were not shown in the table but included in the models. Additionally, as in the previous table, Set size and Quantifier were included in both models. However, as the magnitudes of significant levels were the same across both models for every participant, these variables were left out. For the ‘ANOVA’ part of the table, Significance represents better model fit with Amplitude than without Amplitude, ‘ * ’ for $p < 0.05$; ‘ ** ’ for $p < 0.01$; ‘ *** ’ for $p < 0.001$.

Figure 17

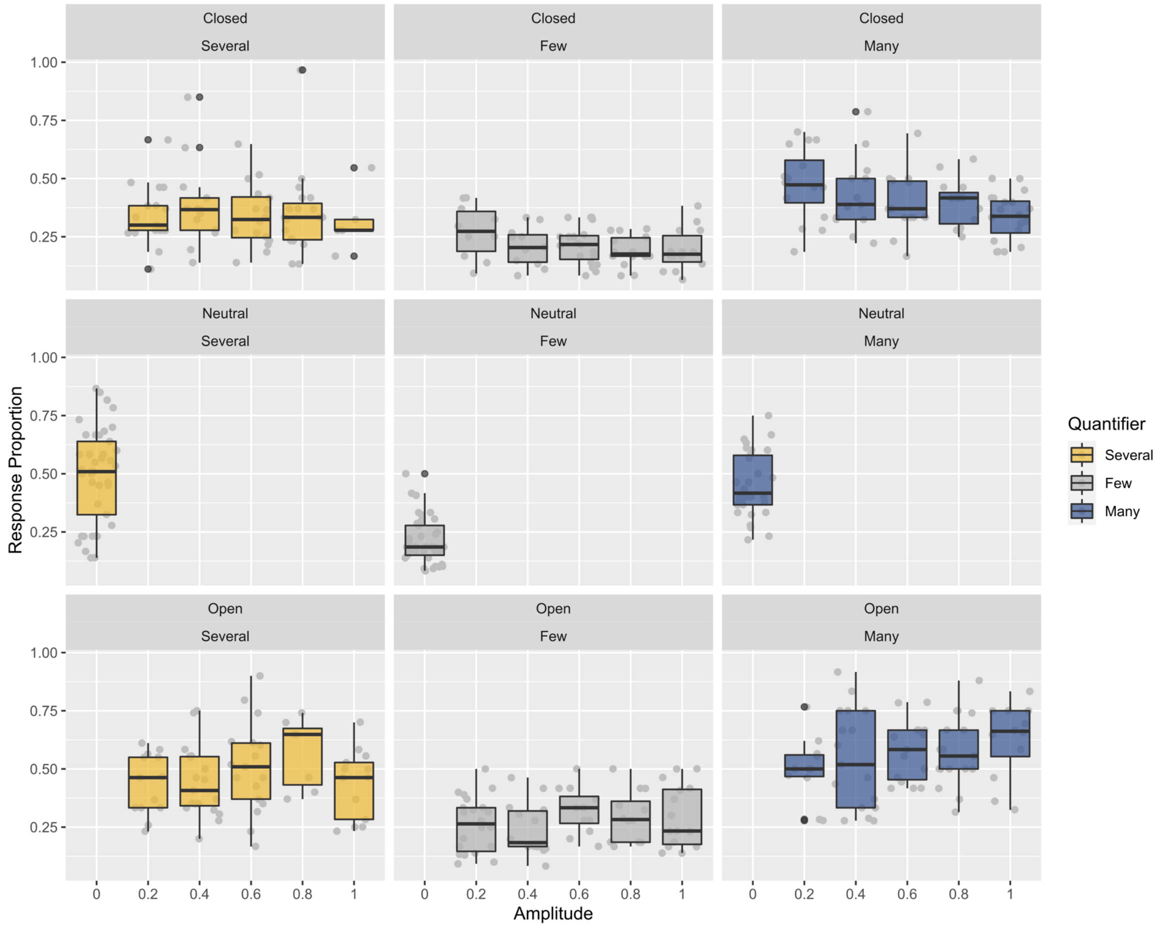
Exemplar Participant's (Subject 4) Number Responses Under Different Amplitudes Subset by Facial Expression Category and Quantifier for Participants in Effects All Explained by the Interaction Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the amplitude categories: from left to right, 0, 0.2, 0.4, 0.6, 0.8 and 1.0 (only the neutral face condition has amplitude 0). Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different facial expression categories and quantifiers: from up to down closing, neutral and opening facial expressions; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Figure 18

Exemplar Participant's (Subject 9) Number Responses Under Different Amplitudes Subset by Facial Expression Category and Quantifier for Participants in Effects All Explained by the Interaction Group

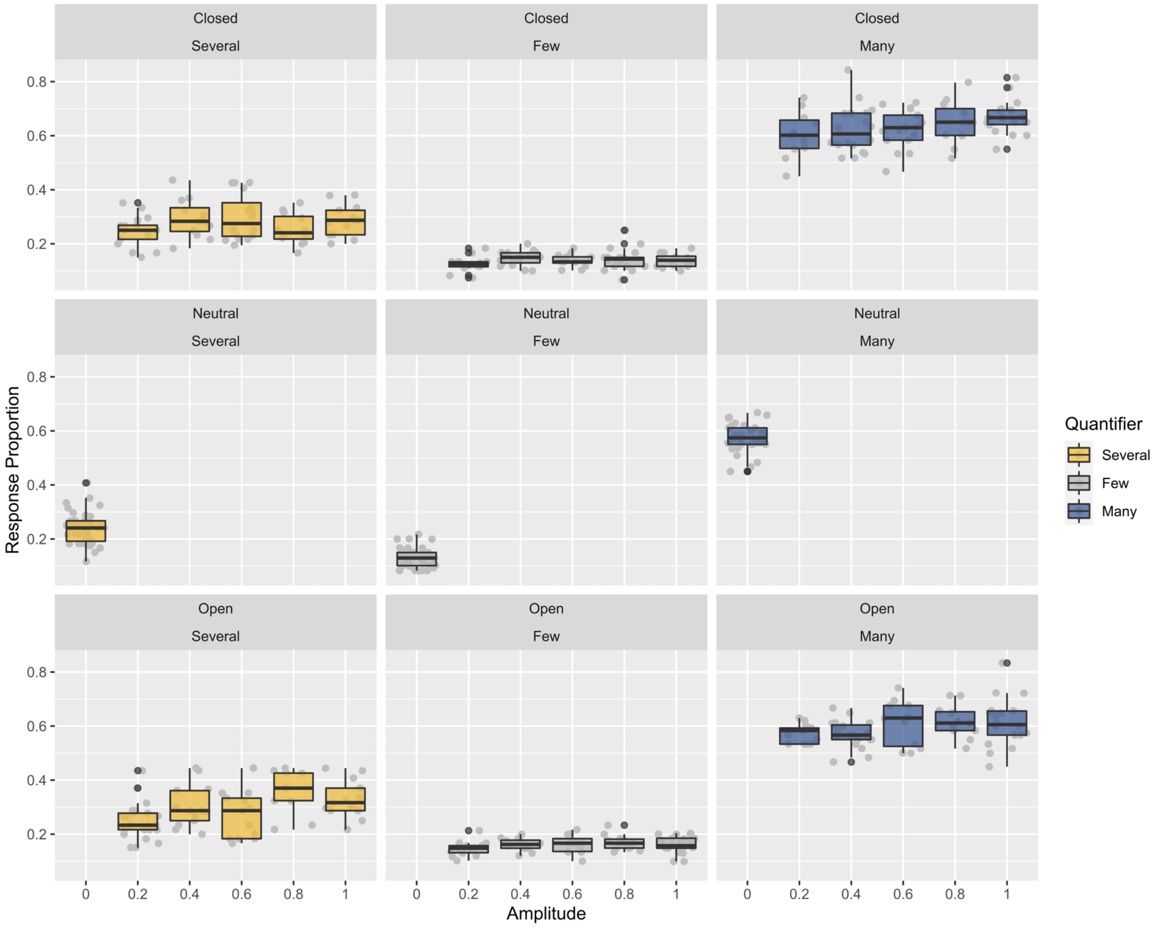


Note. Colourful boxplots and grey colour raw data. The X-axis indicates the amplitude categories: from left to right, 0, 0.2, 0.4, 0.6, 0.8 and 1.0 (only the neutral face condition has amplitude 0). Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different facial expression categories and quantifiers: from up to down closing, neutral and opening facial expressions; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

For 4 participants, including amplitudes showed a significant interaction term but no significant main effect of facial expressions (see Figure S2). For Iconic participants (Subject 4 and Subject 9), for opening facial expressions, the higher the amplitude was, the higher their number response would be; and for closing facial expressions, the higher the amplitude was, the lower their number response would be (Figure 17 and 18). For the Emphasiser participant (Subject 2) and for both the opening and closing facial expressions, higher amplitudes resulted in higher responses (Figure 19). For the opening-only participant (Subject 10) and only for opening facial expressions, higher amplitudes resulted in higher number responses. Interestingly, for Subject 10, the opening facial expressions showed the most apparent effect pattern for the quantifier '*many*', followed by the quantifier '*few*', and finally, the effect was least obvious for the quantifier '*several*' (Figure 20).

Figure 19

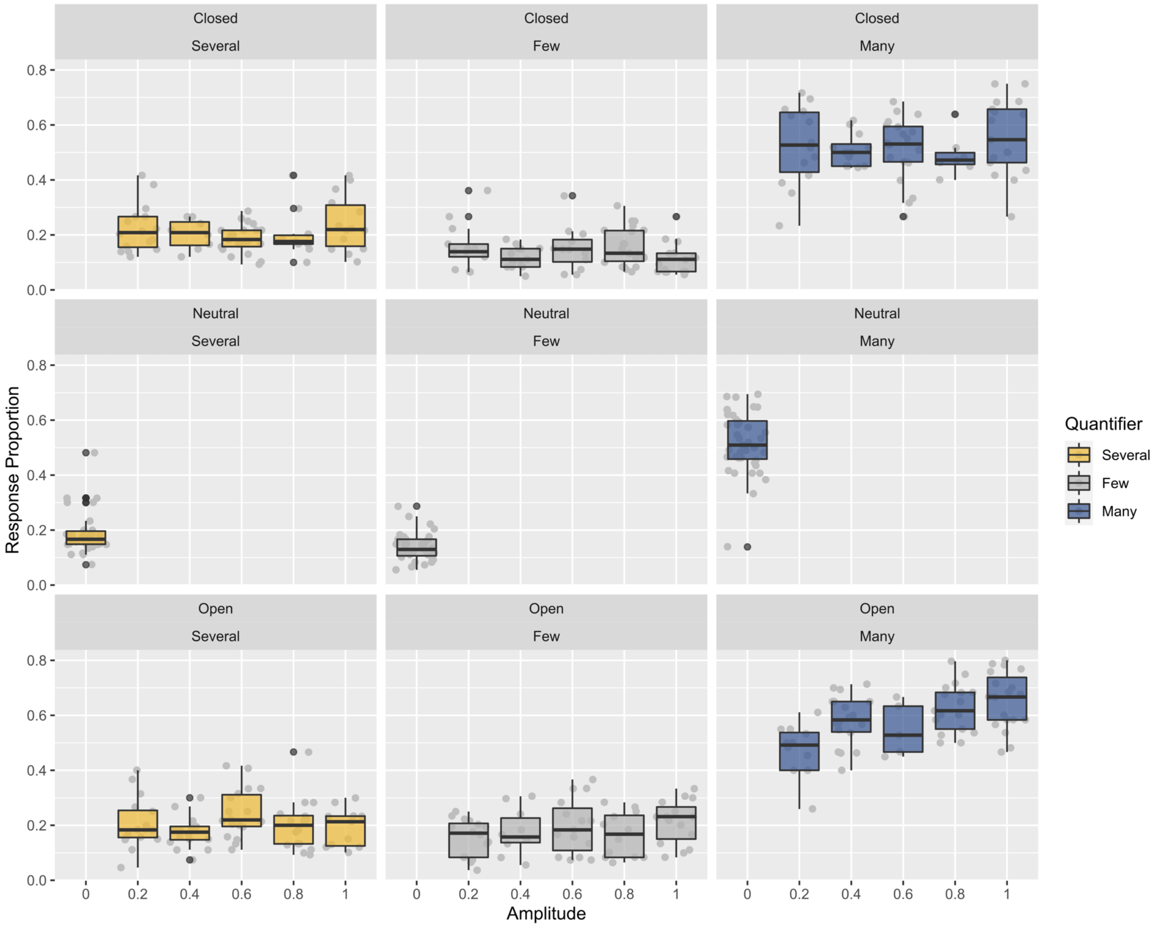
Exemplar Participant's (Subject 2) Number Responses Under Different Amplitudes Subset by Facial Expression Category and Quantifier for Participants in Effects All Explained by the Interaction Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the amplitude categories: from left to right, 0, 0.2, 0.4, 0.6, 0.8 and 1.0 (only the neutral face condition has amplitude 0). Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different facial expression categories and quantifiers: from up to down closing, neutral and opening facial expressions; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Figure 20

Exemplar Participant's (Subject 10) Number Responses Under Different Amplitudes Subset by Facial Expression Category and Quantifier for Participants in Effects All Explained by the Interaction Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the amplitude categories: from left to right, 0, 0.2, 0.4, 0.6, 0.8 and 1.0 (only the neutral face condition has amplitude 0). Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different facial expression categories and quantifiers: from up to down closing, neutral and opening facial expressions; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

As shown in Table 9, for Subject 2, Subject 4, Subject 9 and Subject 10, only the interaction terms were significant: Subject 2 (Emphasiser) closing × Amp: $B_{diff} = 0.042$, SE = 0.014, $t(530.42) = 3.070$, $p = 0.0023$, $d = 0.13$, and opening × Amp: $B_{diff} = 0.055$, SE = 0.013, $t(530.18) = 4.31$, $p = 1.97e\text{-}5$, $d = 0.19$; Subject 4 (Iconic) closing × Amp: $B_{diff} = -0.059$, SE = 0.0081, $t(531.93) = -7.38$, $p = 6.23e\text{-}13$, $d = -0.32$, and opening × Amp: $B_{diff} = 0.054$, SE = 0.0081, $t(530.52) = 6.68$, $p = 5.96e\text{-}11$, $d = 0.29$; Subject 9 (Iconic) closing × Amp: $B_{diff} = -0.096$, SE = 0.031, $t(530.013) = -3.13$, $p = 0.0018$, $d = -0.14$, andopeningg × Amp: $B_{diff} = 0.094$, SE = 0.030, $t(530.019) = 3.15$, $p = 0.0017$, $d = 0.14$; Subject 10 openingg only) opening × Amp: $B_{diff} = 0.090$, SE = 0.021, $t(530.95) = 4.23$, $p =$ 2.82e-5, $d = 0.18$.

Table 9

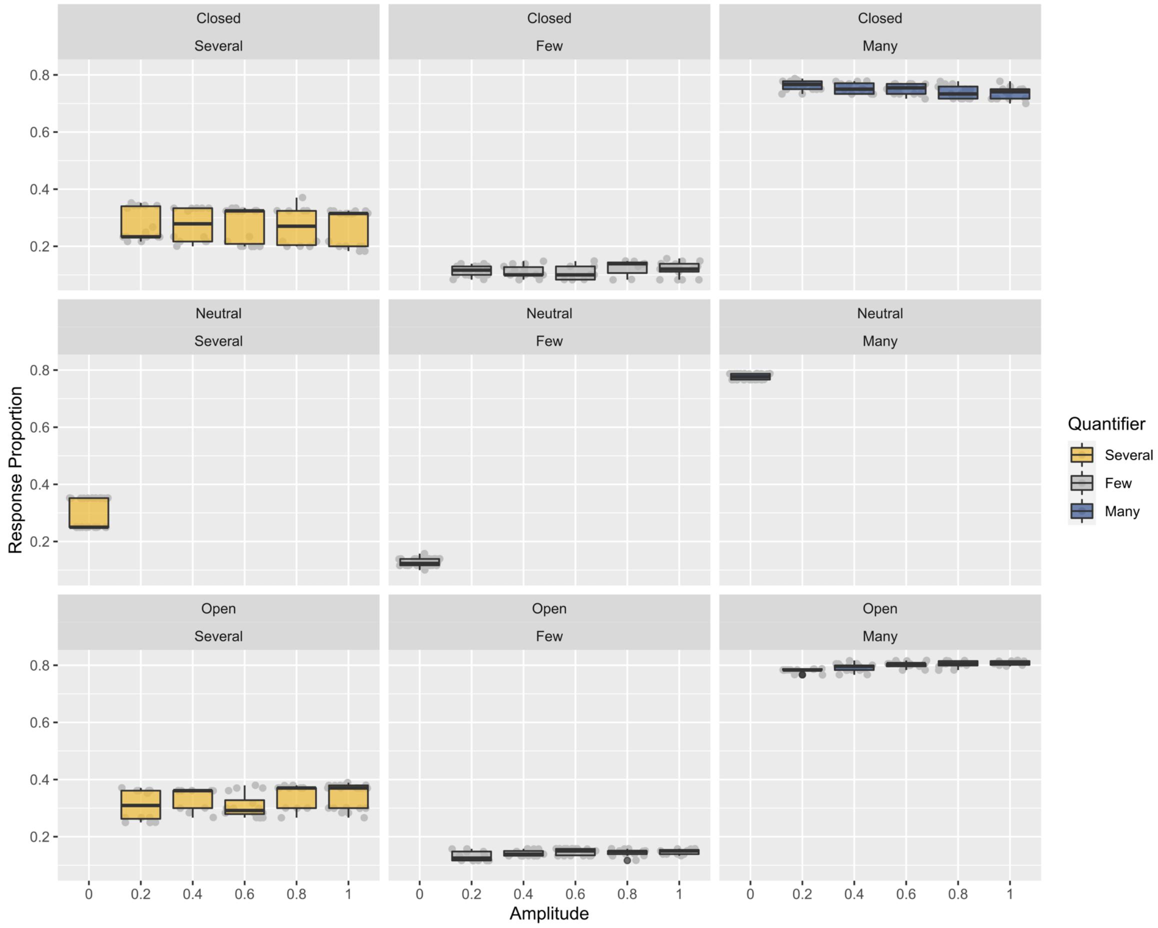
Estimates of Closing and Opening Effects with Their Interactions with Amplitude Achieved from the Linear Mixed Model for Each Participant in Effects All Explained by the Interaction Group

	Estimate	**Std Error**	**df**	**Estimate**	**Std Error**	**df**
		Closing			**Closing:Amp**	
Subject 2	**0.011**	0.011	530.47	**0.042****	0.014	530.42
Subject 4	**0.0028**	0.0062	531.78	**− 0.059*****	0.0081	531.93
Subject 9	**− 0.016**	0.023	530.04	**− 0.096****	0.031	530.01
Subject 10	**0.0030**	0.017	530.78	**0.0098**	0.021	531.57
		Opening			**Opening:Amp**	
Subject 2	**0.0029**	0.010	530.31	**0.055*****	0.013	530.18
Subject 4	**− 0.0064**	0.0062	530.16	**0.054*****	0.0081	530.52
Subject 9	**− 0.018**	0.023	530.03	**0.094****	0.030	530.02
Subject 10	**− 0.011**	0.017	531.04	**0.090*****	0.021	530.95

Note. * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$.

Figure 21

Exemplar Participant's (Subject 3) Number Responses Under Different Amplitudes Subset by Facial Expression Category and Quantifier for Participants in Effects Partly Explained by the Interaction Group

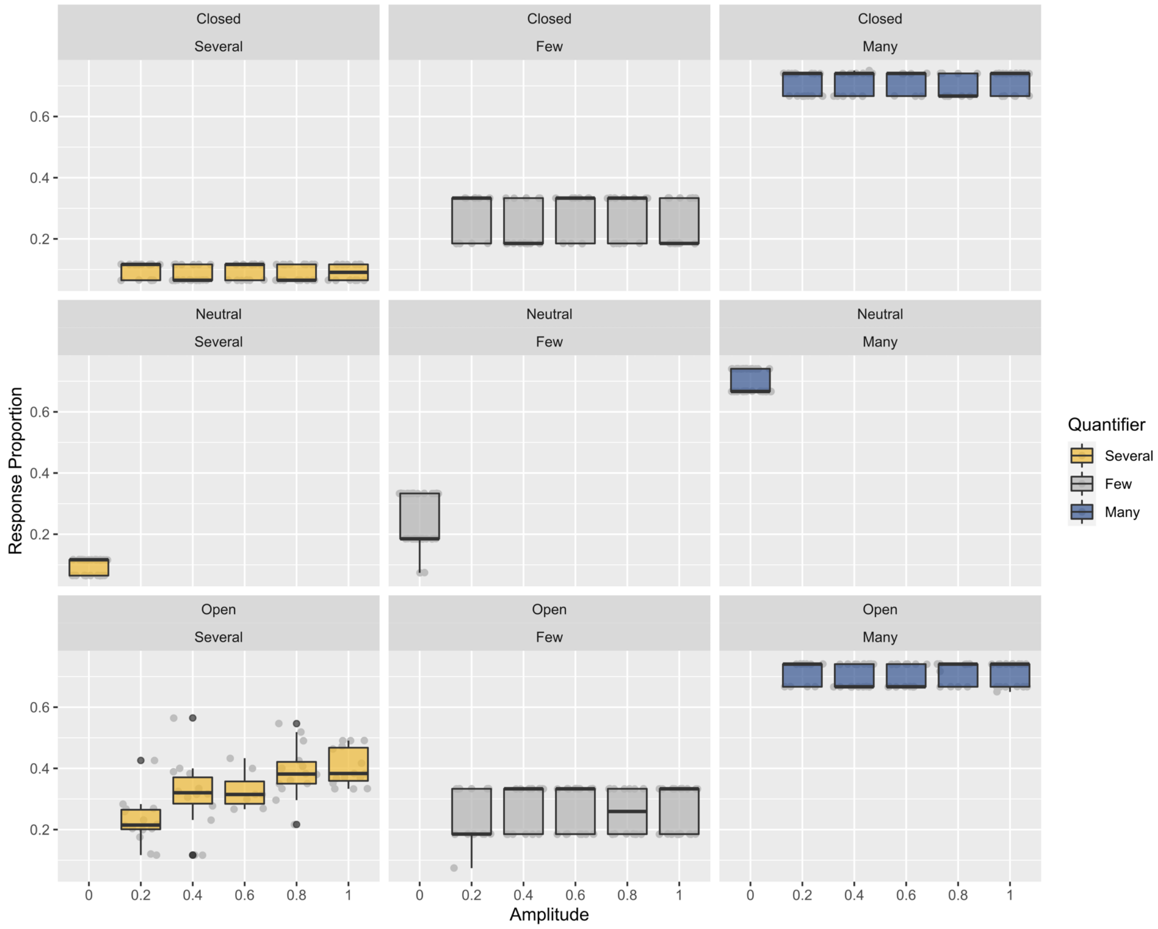


Note. Colourful boxplots and grey colour raw data. The X-axis indicates the amplitude categories: from left to right, 0, 0.2, 0.4, 0.6, 0.8 and 1.0 (only the neutral face condition has amplitude 0). Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different facial expression categories and quantifiers: from up to down closing, neutral and opening facial expressions; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Figure 21 and 22 shows the amplitude effect for participants who had significant interaction terms and some significant main effects of facial expressions. For the iconicity participant (Subject 3), for opening facial expressions, the higher the amplitude was, the higher their number response would be. For closing facial expressions, the higher the amplitude was, the lower their number response would be (Figure 21). However, regarding the main effect of facial expression type, it was not clear. Notably, the within-boxplot variance for quantifier '*several*' was larger than the other boxplots. For the opening-only participant (Subject 8) and only for opening facial expressions, the higher the amplitude was, the higher their number response would be (Figure 22). Interestingly, for Subject 8, the opening facial expression effect was only shown for the quantifier '*several*'; for other boxplots, the number response variances were driven by the set sizes (60 and 108). For the main effect of the opening facial expression, it was clear for the quantifier '*several*' panels (see the difference between Amp 0.0 and Amp 0.2 for opening facial expression).

Figure 22

Exemplar Participant's (Subject 8) Number Responses Under Different Amplitudes Subset by Facial Expression Category and Quantifier for Participants in Effects Partly Explained by the Interaction Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the amplitude categories: from left to right, 0, 0.2, 0.4, 0.6, 0.8 and 1.0 (only the neutral face condition has amplitude 0). Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different facial expression categories and quantifiers: from up to down closing, neutral and opening facial expressions; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Table 10 shows similar funding as Figure 15. For Subject 3 (Iconicity), only the effect of opening facial expression, instead of both opening and closing ones, remained significant (though of smaller effect size) after including the amplitude opening: $M_{diff} = 0.0088$, SE = 0.0043, $t(532) = 2.058$, $p = 0.040$, $d = 0.089$; closing × Amp: $B_{diff} = -0.028$, SE = 0.0051, $t(532) = -5.53$, $p = 5.12e\text{-}8$, $d = -0.24$; opening × Amp: $B_{diff} = 0.020$, SE = 0.0056, $t(532) = 3.64$, $p = 3.02e\text{-}4$, $d = 0.16$). For Subject 8 opening only), the facial expression effect remained significant opening: $M_{diff} = 0.042$, SE = 0.015, $t(532) = 2.72$, $p = 0.0067$, $d = 0.12$; opening × Amp: $B_{diff} = 0.065$, SE = 0.019, $t(532) = 3.42$, $p = 6.79e\text{-}4$, $d = 0.15$).

Table 10

Estimates of Closing and Opening Effects with Their Interactions with Amplitude Achieved from the Linear Mixed Model for Each Participant in Effects Partly Explained by the Interaction Group

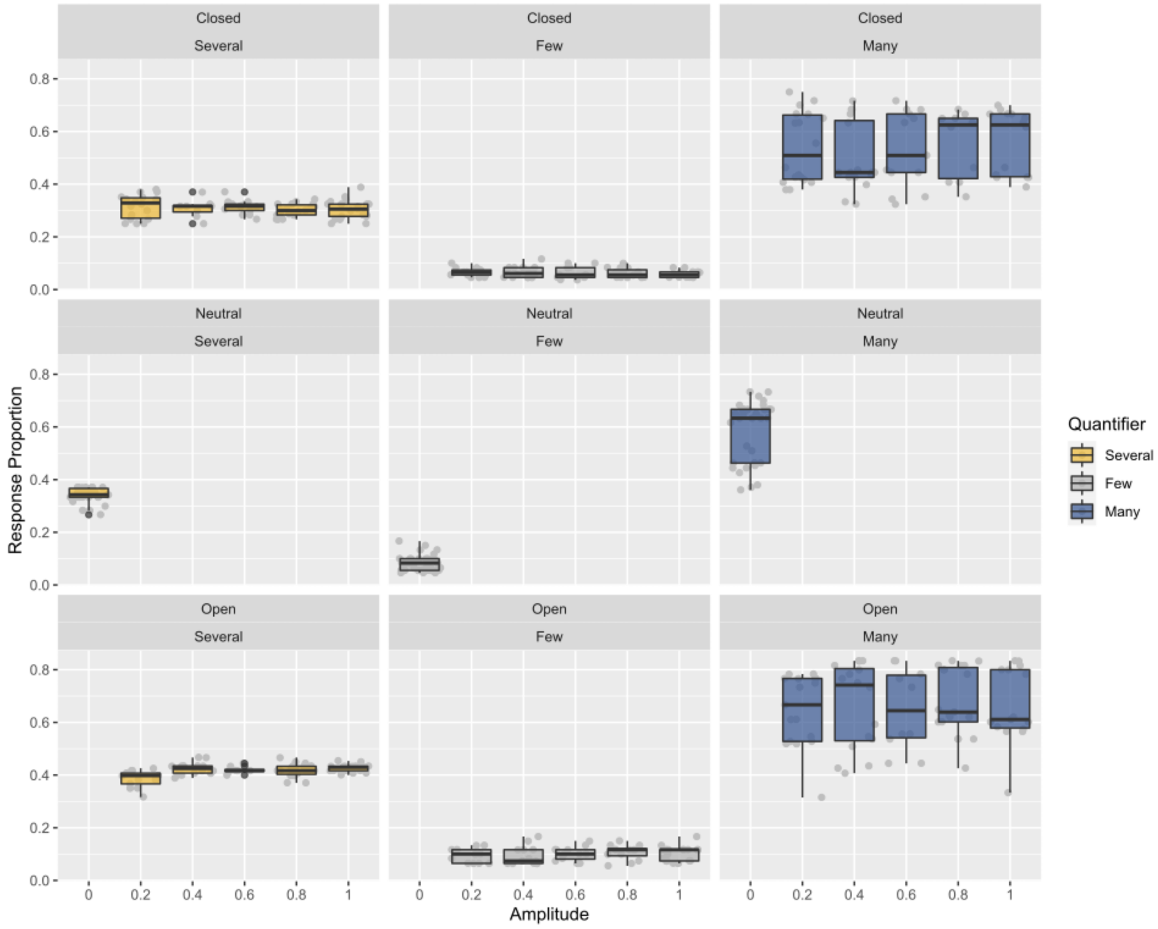
	Estimate	**Std Error**	**df**	**Estimate**	**Std Error**	**df**
		Closing			**Closing:Amp**	
Subject 3	**− 0.0051**	0.004	532	**− 0.028*** **	0.0051	532
Subject 8	**0.0040**	0.015	532	**− 0.0061**	0.019	532
		Opening			**Opening:Amp**	
Subject 3	**0.0088***	0.0043	532	**0.020*****	0.0056	532
Subject 8	**0.042****	0.015	532	**0.065*****	0.019	532

Note. * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$.

Figure 23-24 show the amplitude effect for participants who had no significant interaction terms but significant main facial expressions effects. For the Iconic participant (Subject 1), the amplitude did not play a clear role (Figure 23). Notably, the within-boxplot variance of boxplots with the quantifier '*many*' was visibly larger than the other boxplots. For the Emphasiser participant (Subject 6), similarly, no clear pattern of amplitude effect was found (Figure 24). However, for the closing facial expression × *many* panel, it had a positive amplitude effect. Additionally, the opening facial expression × *several* panel had the largest within-boxplot variance across all boxplots.

Figure 23

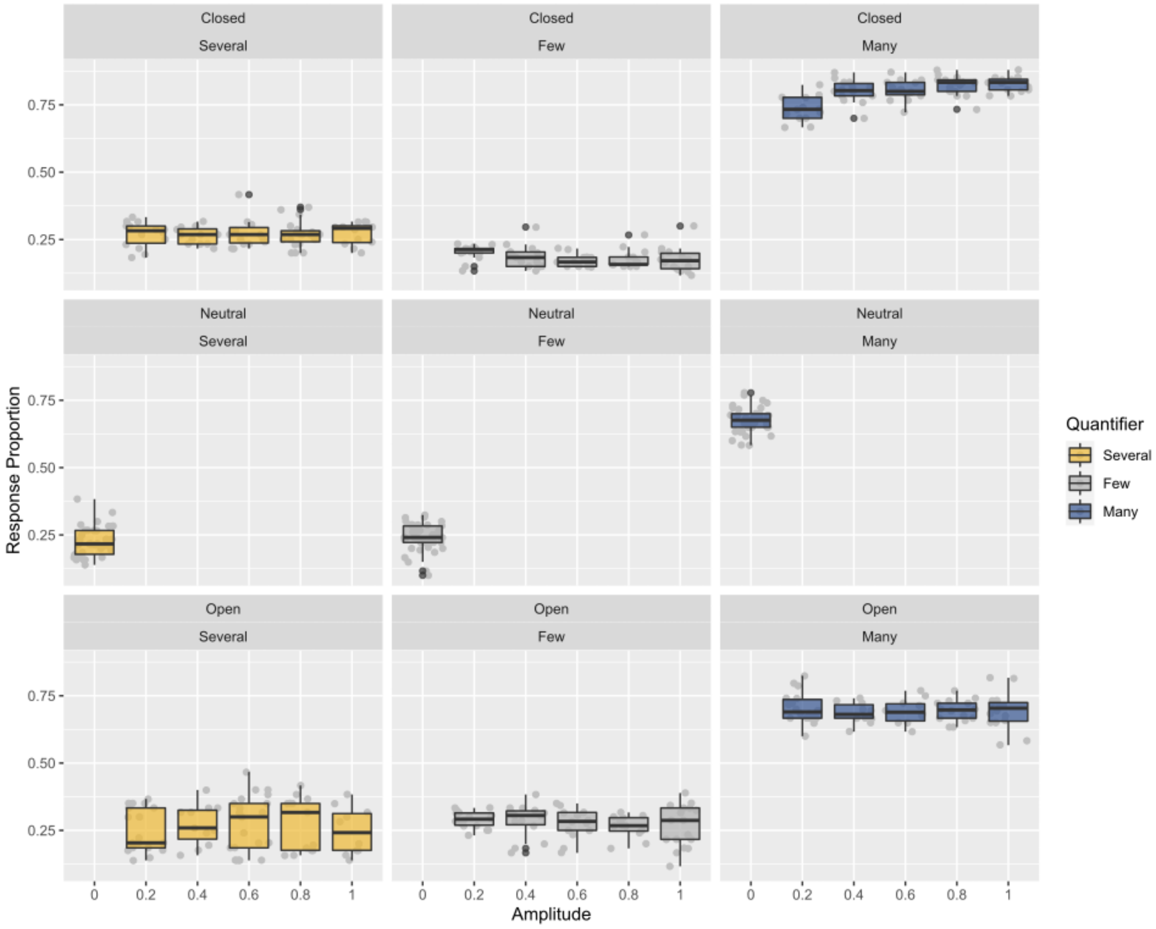
Exemplar Participant's (Subject 1) Number Responses Under Different Amplitudes Subset by Facial Expression Category and Quantifier for Participants in Effects not Explained by the Interaction Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the amplitude categories: from left to right, 0, 0.2, 0.4, 0.6, 0.8 and 1.0 (only the neutral face condition has amplitude 0). Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different facial expression categories and quantifiers: from up to down closing, neutral and opening facial expressions; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Figure 24

Exemplar Participant's (Subject 6) Number Responses Under Different Amplitudes Subset by Facial Expression Category and Quantifier for Participants in Effects not Explained by the Interaction Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the amplitude categories: from left to right, 0, 0.2, 0.4, 0.6, 0.8 and 1.0 (only the neutral face condition has amplitude 0). Y-axis indicates the transformed participant's number responses: Raw responses were divided by their set size. Different panels represent different facial expression categories and quantifiers: from up to down closing, neutral and opening facial expressions; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Table 11

Estimates of closing and opening Effects with Their Interactions with Amplitude Achieved from the Linear Mixed Model for Each Participant in No Amplitude Influence Group

	Estimate	**Std Error**	**df**	**Estimate**	**Std Error**	**df**
		Closing			**Closing:Amp**	
Subject 1	**− 0.032***	0.013	532	**0.012**	0.016	532
Subject 6	**0.019**	0.013	532	**0.032**	0.017	532
		Opening			**Opening:Amp**	
Subject 1	**0.048*****	0.013	532	**0.021**	0.016	532
Subject 6	**0.039****	0.012	532	**− 0.014**	0.016	532

Note. * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$.

Table 11 showed that including the amplitude decreased the main effects of facial expressions while there was no significant interaction. To be specific for Subject 1 (Iconic), both the closing and opening facial expression effect sizes became smaller but still significant, (closing: $M_{diff} = -0.032$, SE = 0.013, $t(532) = -2.52$, $p = 0.012$, $d = -0.11$; opening: $M_{diff} = 0.048$, SE = 0.013, $t(532) = 3.65$, $p = 2.94e\text{-}4$, $d = 0.16$); for Subject 6 (Emphasiser), both the effects became smaller, with the closing facial expression shifted to insignificant (closing: $M_{diff} = 0.019$, SE = 0.013, $t(532) = 1.46$, $p = 0.14$, $d = 0.063$; opening: $M_{diff} = 0.039$, SE = 0.012, $t(532) = 3.18$, $p = 0.0016$, $d = 0.14$).

5 Discussion

Focusing on the perceiver of multimodal spoken messages, here we investigated how people estimate the number indicated by a vague quantifier in the utterance when it is accompanied by facial expressions. We defined opening facial expression, closing facial expression and neutral face. If facial expressions possess iconic potential similar to manual gestures, then opening facial expression should increase the participants' number responses and closing the facial expression should decrease them (compared to neutral face). On the other hand, if facial expressions on the quantifier function more like a generic pragmatic emphasizer similar to stress in the utterance prosody, then both opening and closing facial expressions can have similar functions. We conducted both between-participant analysis and within-participant analysis to investigate the question. For each type of analysis and each participant, we built linear mixed models to examine the effect of Facial Expression

Type and other variables on the participants' number responses. We summarise the main results below and discuss their implications.

5.1 Between-Participant Analysis

In general, the between-participant result showed a significant effect of the opening facial expression but not of the closing facial expression. However, as there was much variation between participants, each of whom might interpret the multimodal messages in their own way, we focused on within-participant analysis to investigate the facial expression effects for each participant.

5.2 Within-Participant Analysis

Within-participant analysis showed that there were 4 of the 10 participants showing the pattern of iconicity. However, these four participants differed in (1) their effect sizes of either the opening or/and closing facial expressions and (2) their variance of responses within a condition (see Figure S1). For example, Subject 9 had a larger within Set Size × Facial Expression Type variance than other participants in this group which possibly derived from the variant factor. The results suggest that the strength of the effect that was consistent among these participants might vary between individuals.

Two participants followed the emphasizer pattern: both the opening and closing facial expressions increased the responses. However, the reason both faces increased the number but not decreased is unclear. It is possible that they still interpreted the presence of a generic facial expression as such as indicating a larger number, so still representing something about the quantity, but less motivated by a direct iconic mapping in both directions.

Additionally, one of the participants (Subject 6; See Figure 14) within this type mentioned during their debriefing that when seeing a face with the closing facial expression, they empathized with the face more, so they were more willing to believe the face. When speaking of *few*, they believed that the closing facial expression meant *really few*, leading to a smaller number; when speaking of *many*, they believed the closing facial expression meant *really many*, leading to a larger number. On the other hand, if it was the opening facial expression, they would think the faces were exaggerating, hence less likely to believe in the faces, leading to higher numbers for *few*. The pattern of '*several*' and '*many*' was less clear. Facial expression for this participant was interpreted as an intensifier modifying the quantity (similar to *really* or *very* in spoken language). These could reflect the influence of emotion signals conveyed by the opening and closing facial expressions. Specifically, as mentioned earlier, the closing facial expression, which contained Brow Lowerer (AU4), is physically similar to facial expressions of emotions with negative valence such as anger and rejection (e.g., Ekman, 1993; Jack et al., 2016; Matsumoto et al., 2008). Hence, Subject 6 might empathize with faces showing the closing facial expression because they perceived the negative emotions. The opening facial expression, which contained Brow Raiser (AU1-2), is physically similar to a surprised facial expression (e.g., Ekman, 1993; Jack et al., 2016; Matsumoto et al., 2008). Intuitively, a surprised facial expression might make the participant think that the number is surprisingly high or surprisingly low—for example, 'few' combined with the opening facial expression could lead to a lower number estimation because the speaker is surprised by how few the objects are. However, this participant did the opposite. One potential explanation is that the participant perceived the facial expression as displaying exaggeration, which attenuated their responses to take this into account.

Two participants followed an opening-face-only pattern: only the opening facial expression increased the responses. Interestingly, as can be seen in Figure 15, for Subject 8, the opening facial expression effect was only present for the quantifier *several*, with slight variance in other boxplots. The result indicates that for this participant *many* and *few* had a fixed proportion (though depending on set size) they represented, which was not influenced by facial expressions and other factors involved in the experiment. In other words, only *several* was interpreted as a quantifier 'vague' enough to vary its meaning based on other factors (in this case, the opening facial expression). Finally, two participants were not influenced by either the opening or closing facial expressions and appeared to disregard the facial expressions.

In conclusion, iconicity was the most prevalent pattern across these 10 participants, while the prevalence of each other patterns—i.e., emphasizer, opening-face-only and no effect—was the same. Comparing the types of facial expressions, 8 out of 10 participants showed a significant effect of the opening facial expression. When the effect was significant, it always increased the responses. On the other hand, 6 out of 10 participants showed a significant effect of the closing facial expression. Interestingly, the significant effect for close faces always coincided with a significant effect of the opening facial expression effect, suggesting that these participants perceived the facial expressions less as a general pragmatic marker and more as depicting quantity iconically. Within the 6 participants that showed a significant effect for close faces, 4 of them showed a significant negative effect (iconicity), and 2 of them showed a positive one (emphasizer). Additionally, 2 participants out of the 10 participants were special in their way of understanding the content of the task. Instead of considering the task as a general conversation, Subject 6 understood the faces specifically in an emotion-communication way. In other words, they tried to understand the emotional state conveyed by the faces and then made the number estimation based on the relationship between the assumed emotional state and the utterance. Subject 8 also differed from other participants in that their responses showed a clear distinction (regarding their average number responses for each quantifier and the variance of their number responses) between quantifier '*several*' and others.

5.3 Amplitudes Influence

To further assess how facial expressions can influence the responses, we further conducted amplitude analyses. We performed analyses for both the between- and within-participant data. For each dataset, we investigated how the model fit changed when including the amplitude factor.

For between-participant analyses, including the amplitude did significantly improve the model fit, and the effect of closing facial expression turned significant because of it.

For within-participant analyses, including the amplitude in the linear model significantly improved the model fit for 6 participants. All these 6 participants were from the eight participants who had at least one significant facial expressions effect, resulting in only two participants that did not significantly respond to the varying strengths of the facial expressions. For the 6 participants, as Figure 17-24 show, their amplitude effect patterns fit well with the expected direction. An interesting finding was from Subject 8: similar to the finding discussed above, the amplitude had a clear pattern only for the quantifier *several* with the opening facial expression. The facial expression effect in *several* explained the high within-boxplot variance observed in the *several* × opening facial expression panel.

In conclusion, the amplitude did influence many of the participants' number responses as expected in a linear fashion—i.e., when the facial expressions increase participants' number responses, the higher the amplitude is, the higher their number response will be; on the other hand, when the facial expressions decrease their responses, the higher the amplitude is, the lower their number response will be.

5.4 Limitations and Future Studies

Our study has several limitations. First, our study had a systematic statistical error: the factor ‘facial expression type’ was correlated with the variable ‘amplitude’. Specifically, the amplitudes varied across [0.2, 0.4, 0.6, 0.8, 1.0] for the opening and closing facial expressions. However, for the neutral face, the only amplitude was 0, as there was no facial expression being presented. This error masked the effect of facial expression type when including the amplitude in the model. Second, we defined the facial expressions in a restricted way. Specifically, we only defined one opening facial expression and one closing facial expression. Future studies can free up the limitation by, for example, allowing different individual facial movements to be combined randomly to investigate how different facial movements (AUs)—per se and in different combinations—can influence number estimation. Third, we did not include the interaction between the quantifier and facial expression type in the model because we anticipated that the facial expressions would influence participant responses in a similar way across different scenarios and quantifiers. However, our more recent data, collected after the completion of this thesis, shows a possible interaction. Specifically, for some participants, both opening and closing facial expressions functioned in largely the same way depending on the quantifiers. With the quantifier ‘few’, the number response average to both opening and closing facial expressions was significantly lower than the response average to neutral faces (no facial expression). With the quantifier ‘many’, the number response average to both opening and closing facial expressions was significantly higher than the response average to neutral faces. We anticipate including this interaction in future analyses and using better models, such as the brms package in R, rather than a linear mixed model.

Our study has several advantages. Different from behavioural studies, which have few trials per participant and follow a between-participant approach, our study had 540 trials per participant. We followed a mixed approach where we conducted both between- and within-participant analyses (resulting in 5400 trials in total for all participants). This approach captures the impact of facial expressions on each participant, summarises the prevalence of effects between-participant, and identifies the range of pragmatic functions of the expressions in the quantity context.

Our study opens new possibilities to further investigate multimodal linguistic phenomena, including estimating the quantity of objects. Using a similar methodology, we can investigate how different modalities—i.e., voice and facial expressions—contribute to communication, including independently (i.e., additive), redundantly, or interacting non-linearly. Adding hand/body gestures would enable further investigation of how vocal-facial-gestural cues interact. Second, studies can, in a multimodal setting, investigate the iconic facial signals that indicate acceptance and rejection (see the preliminary work: Nölle et al., 2021). Third, we anticipate investigating other linguistic phenomena, such as the perceived confidence or doubt of a speaker, referred to as displaying the "feeling of other's knowing" (Brennan & Williams, 1995), the degree of (im)politeness (Brown & Levinson, 1987) and appropriateness of apologies (George et al., 2022).

In summary, our work underscores the increasing agreement in the field that language should be studied within a broader multimodal context to reflect the environment in which it has developed—i.e., face-to-face conversations (Holler & Levinson, 2019). Our recent work further highlights the importance of doing so—for example, we found that Brow Raiser is perceived as indicating greater confidence in speakers, which mirrors their use in signed languages to emulate raised pitch as in spoken language (Mapson, 2014). By further examining how facial expressions contribute to multimodal communication, including with different linguistic phenomena, we can better understand their different functions and flexibility as a tool for social interaction.

6 Conclusions

Here, we show that facial expressions can pragmatically represent quantities in an iconic way. Overall, some participants interpreted speakers displaying opening facial movements as referring to higher numbers and those displaying closing facial movements as referring to lower numbers. We also observed individual variance where both the opening and closing facial movements are interpreted as emphasizers, thereby increasing their number responses regardless of facial expression types. To our knowledge, this is the first study to show that facial expressions can impact the perceived semantic meaning of spoken vague quantifiers. We anticipate that our findings will enable a deeper understanding of multimodal communication that includes facial expressions.

7 References

Ahumada, A. and Lovell, J. (1971) 'Stimulus Features in Signal Detection', *The Journal of the Acoustical Society of America*, 49(6B), pp. 1751–1756. Available at: https://doi.org/10.1121/1.1912577.

Bavelas, J. and Chovil, N. (2018) 'Some pragmatic functions of conversational facial gestures1', *Gesture*, 17(1), pp. 98–127.

Bavelas, J., Gerwing, J. and Healing, S. (2014) *Hand and facial gestures in conversational interaction*. The Oxford handbook of language and social psychology.

Chen, C. *et al.* (2018) 'Distinct facial expressions represent pain and pleasure across cultures', *Proceedings of the National Academy of Sciences*, 115(43). Available at: https://doi.org/10.1073/pnas.1807862115.

Coventry, K.R. *et al.* (2010) 'Talking about quantities in space: Vague quantifiers, context and similarity', *Language and Cognition*, 2(2), pp. 221–241. Available at: https://doi.org/10.1515/langcog.2010.009.

Darwin, C. and Prodger, P. (1998) *The Expression of the Emotions in Man and Animals*. Oxford University Press.

Domaneschi, F., Passarelli, M. and Chiorri, C. (2017) 'Facial expressions and speech acts: experimental evidences on the role of the upper face as an illocutionary force indicating device in language comprehension', *Cognitive Processing*, 18(3), pp. 285–306. Available at: https://doi.org/10.1007/s10339-017-0809-6.

Edward, T. (1966) 'Hall, The Hidden Dimension'. Garden City, NY: Doubleday.

Ekman, P. (1971) 'Universals and cultural differences in facial expressions of emotion', *Nebraska Symposium on Motivation*, 19, pp. 207–283.

Ekman, P. (1993) 'Facial expression and emotion.', *American psychologist*, 48(4), p. 384.

Ekman, P. and Friesen, W.V. (1978) 'Facial action coding system', *Environmental Psychology & Nonverbal Behavior* [Preprint].

Gill, D. *et al.* (2014) 'Facial Movements Strategically Camouflage Involuntary Social Signals of Face Morphology', *Psychological Science*, 25(5), pp. 1079–1086. Available at: https://doi.org/10.1177/0956797614522274.

Hassemer, J. and Winter, B. (2018) 'Decoding Gestural Iconicity', *Cognitive Science*, 42(8), pp. 3034–3049. Available at: https://doi.org/10.1111/cogs.12680.

Haugh, M. and Bousfield, D. (2012) 'Mock impoliteness, jocular mockery and jocular abuse in Australian and British English', *Journal of Pragmatics*, 44(9), pp. 1099–1114. Available at: https://doi.org/10.1016/j.pragma.2012.02.003.

Hensel, L.B. *et al.* (2021) 'Social Trait Facial Expressions Comprise Latent Affective Facial Signals', *Journal of Vision*, 21(9), pp. 1988–1988.

Holler, J. *et al.* (2015) 'Eye'm talking to you: speakers' gaze direction modulates co-speech gesture processing in the right MTG', *Social Cognitive and Affective Neuroscience*, 10(2), pp. 255–261. Available at: https://doi.org/10.1093/scan/nsu047.

Holler, J., Kendrick, K.H. and Levinson, S.C. (2018) 'Processing language in face-to-face conversation: Questions with gestures get faster responses', *Psychonomic Bulletin & Review*, 25(5), pp. 1900–1908. Available at: https://doi.org/10.3758/s13423-017-1363-z.

Holler, J. and Levinson, S.C. (2019) 'Multimodal Language Processing in Human Communication', *Trends in Cognitive Sciences*, 23(8), pp. 639–652. Available at: https://doi.org/10.1016/j.tics.2019.05.006.

Ibrahimagić-Šeper, L. *et al.* (2006) 'Anthropometric differences between males and females in face dimensions and dimensions of central maxillary incisors', *Medicinski glasnik*, 3(2), pp. 58–62.

Ince, R.A.A., Kay, J.W. and Schyns, P.G. (2022) 'Within-participant statistics for cognitive science', *Trends in Cognitive Sciences*, 26(8), pp. 626–630. Available at: https://doi.org/10.1016/j.tics.2022.05.008.

Jack, R.E. *et al.* (2012) 'Facial expressions of emotion are not culturally universal', *Proceedings of the National Academy of Sciences*, 109(19), pp. 7241–7244. Available at: https://doi.org/10.1073/pnas.1200155109.

Jack, R.E. *et al.* (2016) 'Four not six: Revealing culturally common facial expressions of emotion.', *Journal of Experimental Psychology: General*, 145(6), pp. 708–730. Available at: https://doi.org/10.1037/xge0000162.

Jack, R.E. and Schyns, P.G. (2017) 'Toward a Social Psychophysics of Face Communication', *Annual Review of Psychology*, 68(1), pp. 269–297. Available at: https://doi.org/10.1146/annurev-psych-010416-044242.

Lakoff, G. and Johnson, M. (1980) *Metaphors we live by*. University of Chicago press.

Lee, D.H. *et al.* (2014) 'Optical Origins of Opposing Facial Expression Actions', *Psychological Science*, 25(3), pp. 745–752. Available at: https://doi.org/10.1177/0956797613514451.

Matsumoto, D. *et al.* (2008) 'Facial expressions of emotion.'

McGurk, H. and MacDonald, J. (1976) 'Hearing lips and seeing voices', *Nature*, 264(5588), pp. 746–748.

Newstead, S.E., Pollard, P. and Riezebos, D. (1987) 'The effect of set size on the interpretation of quantifiers used in rating scales', *Applied Ergonomics*, 18(3), pp. 178–182. Available at: https://doi.org/10.1016/0003-6870(87)90001-9.

Ortega, G. and Özyürek, A. (2020) 'Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture', *Behavior Research Methods*, 52(1), pp. 51–67. Available at: https://doi.org/10.3758/s13428-019-01204-6.

Partan, S. and Marler, P. (1999) 'Communication Goes Multimodal', *Science*, 283(5406), pp. 1272–1273. Available at: https://doi.org/10.1126/science.283.5406.1272.

Perniss, P. (2018) 'Why We Should Study Multimodal Language', *Frontiers in Psychology*, 9, p. 1109. Available at: https://doi.org/10.3389/fpsyg.2018.01109.

Roberts, J.A., Taylor, P.W. and Uetz, G.W. (2007) 'Consequences of complex signaling: predator detection of multimodal cues', *Behavioral Ecology*, 18(1), pp. 236–240. Available at: https://doi.org/10.1093/beheco/arl079.

Susskind, J.M. *et al.* (2008) 'Expressing fear enhances sensory acquisition', *Nature Neuroscience*, 11(7), pp. 843–850. Available at: https://doi.org/10.1038/nn.2138.

Vigliocco, G., Perniss, P. and Vinson, D. (2014) 'Language as a multimodal phenomenon: implications for language learning, processing and evolution', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), p. 20130292. Available at: https://doi.org/10.1098/rstb.2013.0292.

Winter, B., Marghetis, T. and Matlock, T. (2015) 'Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number', *Cortex*, 64, pp. 209–224. Available at: https://doi.org/10.1016/j.cortex.2014.10.015.

Winter, B., Perlman, M. and Matlock, T. (2013) 'Using space to talk and gesture about numbers: Evidence from the TV News Archive', *Gesture*, 13(3), pp. 377–408. Available at: https://doi.org/10.1075/gest.13.3.06win.

Woodin, G. *et al.* (2020) '"Tiny numbers" are actually tiny: Evidence from gestures in the TV News Archive', *PLOS ONE*. Edited by D. Tay, 15(11), p. e0242142. Available at: https://doi.org/10.1371/journal.pone.0242142.

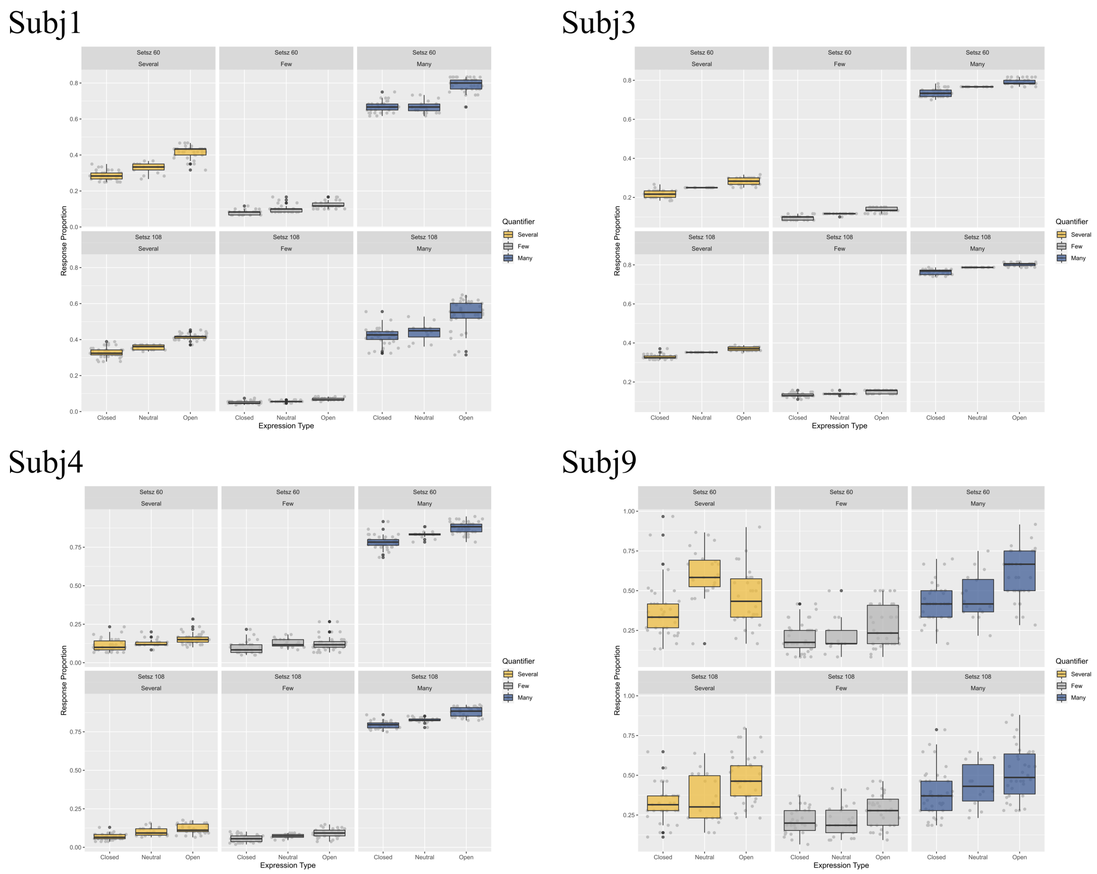
Yu, H., Garrod, O.G. and Schyns, P.G. (2012) 'Perception-driven facial expression synthesis', *Computers & Graphics*, 36(3), pp. 152–162.

Zhan, J. *et al.* (2019) 'Modelling face memory reveals task-generalizable representations', *Nature Human Behaviour*, 3(8), pp. 817–826. Available at: https://doi.org/10.1038/s41562-019-0625-3.

8 Appendices

Figure S1

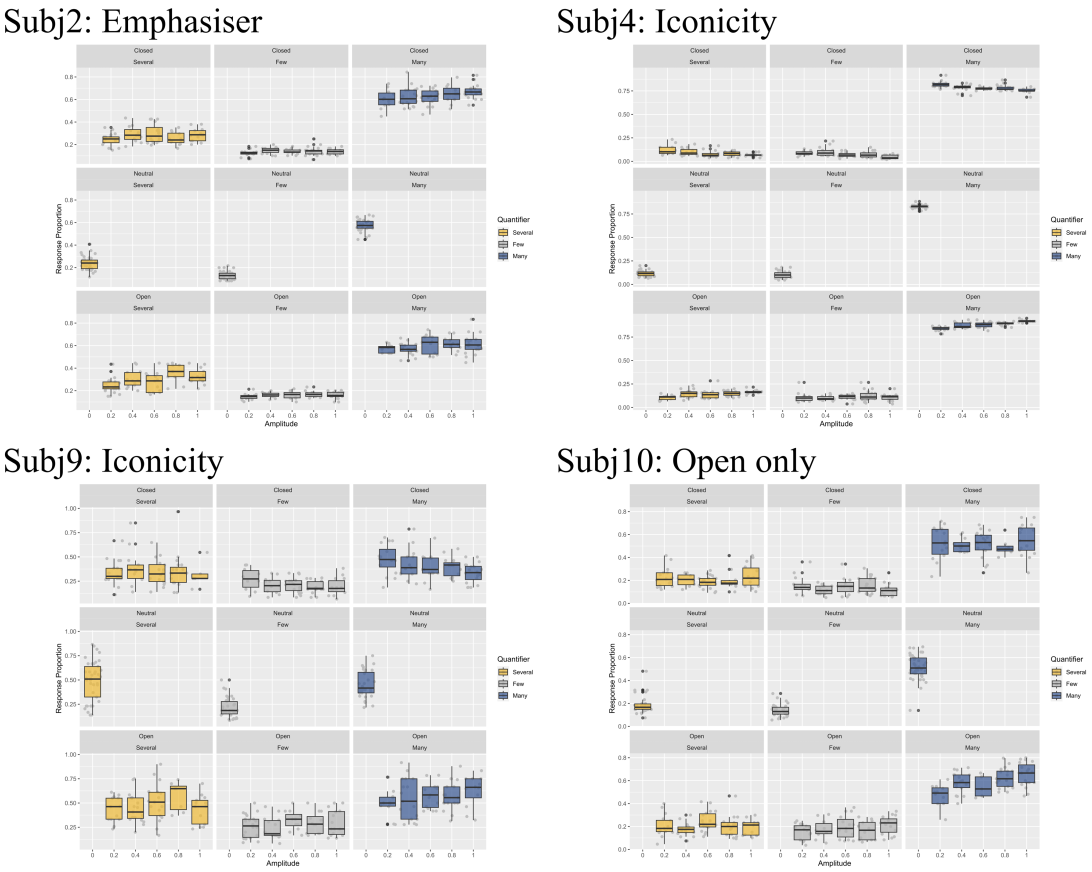
Per Participant's Number Responses Under Different Facial Expressions Subset by Set Size and Quantifier for Participants from Iconic Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the facial expression category: from left to right, closing, Neutral and opening facial expressions. Y axis indicates the transformed per participant's number responses: Raw responses were divided by their set size. Different panels represent different participants. Different sub-panels represent different set sizes and quantifiers: rows for the former, from up to down 60 and 108; columns for the latter, from left to right '*several*', '*few*' and '*many*'.

Figure S2

Per Participant's Number Responses Under Different Amplitudes Subset by Facial Expression Category and Quantifier for Participants in Effects All Explained by the Interaction Group



Note. Colourful boxplots and grey colour raw data. The X-axis indicates the amplitude categories: from left to right, 0, 0.2, 0.4, 0.6, 0.8 and 1.0 (only the neutral face condition has amplitude 0). Y-axis indicates the transformed per participant's number responses: Raw responses were divided by their set size. Different panels represent different participants. Different sub-panels represent different facial expression categories and quantifiers: from up to down closing, neutral and opening facial expressions; columns for the latter, from left to right '*several*', '*few*' and '*many*'.