Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-019-01948-5

Understanding the visual perception of awkward body movements: How interactions go awry

Akila Kadambi[1] · Nicholas Ichien[1] · Shuwen Qiu[2] · Hongjing Lu[1,3]

Abstract
Dyadic interactions can sometimes elicit a disconcerting response from viewers, generating a sense of "awkwardness." Despite the ubiquity of awkward social interactions in daily life, it remains unknown what visual cues signal the oddity of human interactions and yield the subjective impression of awkwardness. In the present experiments, we focused on a range of greeting behaviors (handshake, fist bump, high five) to examine both the inherent objectivity and impact of contextual and kinematic information in the social evaluation of awkwardness. In Experiment 1, participants were asked to discriminate whether greeting behaviors presented in raw videos were awkward or natural, and if judged as awkward, participants provided verbal descriptions regarding the awkward greeting behaviors. Participants showed consensus in judging awkwardness from raw videos, with a high proportion of congruent responses across a range of awkward greeting behaviors. We also found that people used social-related and motor-related words in their descriptions for awkward interactions. Experiment 2 employed advanced computer vision techniques to present the same greeting behaviors in three different display types. All display types preserved kinematic information, but varied contextual information: (1) *patch displays* presented blurred scenes composed of patches; (2) *body displays* presented human body figures on a black background; and (3) *skeleton displays* presented skeletal figures of moving bodies. Participants rated the degree of awkwardness of greeting behaviors. Across display types, participants consistently discriminated awkward and natural greetings, indicating that the kinematics of body movements plays an important role in guiding awkwardness judgments. Multidimensional scaling analysis based on the similarity of awkwardness ratings revealed two primary cues: motor coordination (which accounted for most of the variability in awkwardness judgments) and social coordination. We conclude that the perception of awkwardness, while primarily inferred on the basis of kinematic information, is additionally affected by the perceived social coordination underlying human greeting behaviors.

Keywords Action perception · Human interactions · Social perception · Awkwardness

In some social situations, dyadic interactions can elicit a disconcerting response from viewers—a sense of "awkwardness." For example, recent videos of U.S. President Donald Trump shaking hands with his Supreme Court nominee Neil Gorsuch (link) and the Prime Minister of Japan (link) generated a wave of discussions among both laypeople and experts about what exactly constitutes an "awkward" motoric social interaction (link).

✉ Akila Kadambi
akadambi@ucla.edu

[1] Department of Psychology, UCLA, Los Angeles, CA, USA

[2] Department of Computer Sciences, UCLA, Los Angeles, CA, USA

[3] Department of Statistics, UCLA, Los Angeles, CA, USA

Although difficult to pinpoint an exact definition, awkwardness is a subjective impression that can arise from many different cues, including a failure in executing motor behaviors, misunderstood intentions, and conflicting personality traits. While perhaps amusing as a construct, when an interpersonal interaction is perceived as awkward, either by participants or third-party observers, social goals and fluid communication are also likely to be impeded (Snyder, Tanke, & Berscheid, 1977), compromised more readily in clinical disorders associated with atypical mentalizing ability, such as Autism (Heavey, Phillips, Baron-Cohen, & Rutter, 2000). In this regard, awkwardness is a highly complex social behavior as it requires both the sophisticated understanding of social heuristics and the subsequent knowledge of perceived violations. Therefore, to sufficiently study this underexamined, yet complex and heterogenous construct important questions emerge: Are people idiosyncratic in their

Published online: 30 January 2020

perceptions of awkwardness in dyadic interactions, or are people in general agreement? If people agree with each other in perceiving awkwardness, is it possible to pin down the visual characteristics contributing to the impression of awkwardness? What visual features signal the oddity of a dyadic interaction? The present paper explores these questions through examining judgments of awkwardness conveyed through human social greeting interactions in naturalistic videos.

Two research fields provide relevant knowledge about perceiving social attributes from visual input. Research on biological motion perception shows that when human actions are reduced to moving dots located at key joints (Johansson, 1976), human observers make reliable attributions of social properties such as deception (Runeson & Frykholm, 1983), intention (Hohmann, Troje, Olmos, & Munzert, 2011), affect (Pollick, Paterson, Bruderlin, & Sanford, 2001), sex (Johnson & Tassinary, 2005), body identity (Cutting & Kowzlowski, 1977; Burling, Kadambi, Safari, & Lu, 2019), and personality traits (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004). Additionally, these studies highlight the importance of kinematic information in inferring the social properties of human actions.

A separate line of work in person perception has focused on a different but equally important question—the role of visual context in social evaluation. Humans live and interact within rich contextual environments. Human observers have been shown to use static images of faces to make reliable attributions of visually ambiguous social properties such as sexual orientation (Rule & Ambady, 2008; Rule, Ambady, Adams, & Macrae, 2008), political identity (Rule & Ambady, 2010), and personality traits (see Todorov, Said, & Verosky, 2011 for a review). Related research has demonstrated that other aspects of visual context, such as race (Alter, Stern, Granot, & Balcetis, 2016), attire (Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011), and scene background (Freeman, Ma, Han, & Ambady, 2013) also influence the attribution of visually based social properties. From these categorizations, visual context is multifaceted and includes a wide range of cues, from person identity, which contextualizes a face (Freeman & Ambady, 2011), to high-level background information, such as the physical scene and the social environment in which a human is grounded.

Through integrating the two separate but closely linked research fields, we provide a novel methodology to examine the role of visual context and human kinematics of body movements in the perception of awkward greeting behaviors. Modern advances in deep learning models make it possible to systematically segregate motor and contextual information, by segmenting body movements from background scenes in raw action video recordings. To this end, we delineated varying aspects of visual context (described further below), through presenting naturalistic and contextualized human interactions in different display types, to parametrically examine the contribution of contextual and kinematic information in perceived awkwardness.

To assess whether people are in general agreement in their perceptions of awkwardness in dyadic interactions, we used both free response (Experiment 1) and rating (Experiment 2) paradigms. In Experiment 1, participants were presented with a variety of daily encountered greeting behaviors in raw videos and asked to identify whether each greeting behavior appeared awkward. If the video was categorized as awkward, participants were asked to describe why this categorization was made. The text analysis based on free responses allowed us to explore whether people consistently attend to certain social or motor cues in the raw videos to identify the presence of an awkward interaction.

Experiment 2 aimed to further examine the interpretability of the free response results by asking for subjective ratings on the experimentally manipulated stimuli. We employed advanced computer vision techniques to generate the stimuli of greeting actions presented in three different video display types. These displays consisted of dyadic interactions and varied the amount of visual context presented in the stimuli. To parameterize the amount of visual information, we broadly divided visual context into four main categories: body structure (i.e., limb articulations and height), body morphology (i.e., width, body shape, gender), actor identity (i.e., skin tone, coarse facial features, attire), and scene depiction (i.e., physical background—indoor versus outdoor scenes). We characterized kinematics as the information provided by body movements of the actors involved in each greeting action. Across the display types, we maintained the kinematic information of body movements, while removing particular categories of contextual information. For instance, in one type of display (discussed in more detail below), we removed the scene and actor identity information, but maintained body morphology and structure characteristics. In another display type, consisting of the sparsest visual information, we removed the scene, actor identity, and body morphology information, while just preserving the body structure. Through incorporating these categorizations of visual context, we examined the independent contribution of human kinematic information to awkwardness judgments as affected by the gradual mitigation of contextual information in the different display types. To further elaborate, we describe the visual context provided by three different display types (examples are shown in Fig. 1).

The first type, *patch displays*, presented blurred scenes and featured the most amount of visual context of our three display types. Specifically, the patch displays preserved all four components of our broadly defined criteria of visual context: scene depiction, actor identity, body morphology, and body structure. They offered rich cues about the settings in which greeting actions occurred, including the scene background, objects

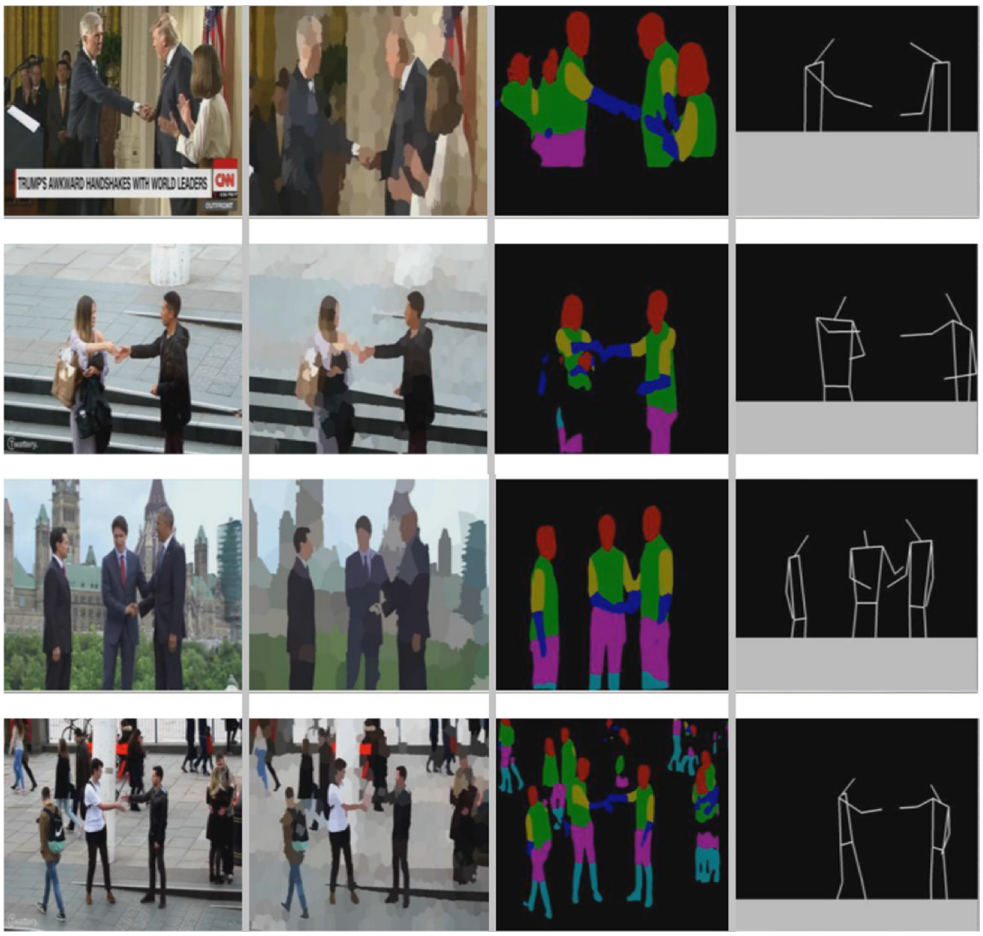


Fig. 1 Example of stimuli. Column 1, raw YouTube video frames; column 2, patch display processed by superpixel algorithm; column 3, body display processed by RefineNet model; column 4, skeleton display converted from raw video

in the scene, and other people not involved in greeting actions. The patch displays also provided cues related to actor identities including skin tone, coarse facial features (e.g., a separation of face area and hair), and attire. Additionally, these displays preserved body morphology such as body shape and gender, as well as structural body information (limb articulations and height).

The second type, *body displays*, presented human body figures with varying colors for different body parts on a black background. The body displays provided less visual context than did patch displays. Specifically, both scene information (the physical environment) and actor identity information (skin tone, coarse facial features, attire) were removed, while preserving body morphology (cues about coarse body shape, such as width and gender) and body structure (joints and height). Note that while the physical scene was eliminated, sparse cues about other actors not involved in greeting behaviors (occasionally displaying body parts of background actors) remained.

The third type, *skeleton displays*, presented white skeleton figures resembling human body structure against a black background. The skeleton displays featured the least visual context of our three display types, preserving only the structural body information of the main actors involved in greeting behaviors. Specifically, the skeleton displays included no cues about the background environment in which greeting actions occurred, nor actor identity information such as skin tone and coarse facial expression, nor cues about body shape, such as body width and gender. Therefore, only body height and limb articulation information was presented, depicted by stick figures of the main actors involved in greeting behaviors.

Notably, all three display types held human kinematics constant, as they were generated from the identical greeting behaviors from the recorded videos, but each display included different categories of visual context. This key experimental manipulation allowed us to compare awkwardness ratings of greeting behaviors across a range of naturalistic actions, in order to examine the relationship between visual context and kinematic information in the social evaluation of greeting behaviors. To underscore, these display types were chosen because of their inherent consistency in the presentation of human body structure and kinematics, while varying key components of visual context: body structure (consistent across displays), body morphology, actor identity, and background scene information.

Finally, Experiment 2 also attempted to explore what key visual features might serve as cues to the perception of awkwardness. To this end, we conducted a multidimensional scaling (MDS) analysis based on participant awkwardness ratings to infer the two-dimensional psychological space for perceived awkwardness. The interpretation of the two primary dimensions could provide connections to the characteristics of word descriptions reported in Experiment 1 when people were explicitly asked to describe awkward greetings.

Experiment 1

A free-response study was first conducted to measure the perception of awkwardness in greeting behaviors from raw video recordings. Here, participants viewed videos of awkward and natural greeting behaviors and subsequently categorized the video as awkward or natural by providing written descriptions of the social interaction. The present experiment also explored features of the semantic descriptions judging awkward interactions, and whether the contribution of social and motor cues indeed signified the presence of an awkward interaction.

Method

Participants

Thirty participants (male = 9, female = 21) were recruited from the University of California, Los Angeles (UCLA) Psychology Subject pool. Participants provided informed consent, as approved by the UCLA Institutional Review Board (#16-001879), and were given course credit for their participation.

Stimuli and apparatus

Participants were tested in a dark, quiet room 76.2 cm from the display. Monitor width and height was $53.1^{o} \times 40.7^{o}$. Thirty-four videos from YouTube (see Appendix A for links) were selected to capture a variety of greeting behaviors ranging from awkward and natural social interactions. The videos varied in length (2 to 27 seconds, $M = 6.36$, $SD = 4.47$) and context and were selected according to two important reasons. First, previous studies on interpersonal coordination of joint actions have largely categorized actions as either deliberate (such as high-fiving) or unnoticed (such as bumping into each other; e.g., Schmidt, Fitzpatrick, Caron, & Mergeche, 2011). By including videos that encompass both types of greeting actions, we were able to test a more variable set of interactions. Secondly, the selected videos depicted different degrees of awkwardness. The range of variability allowed for participants to appraise social situations that were more relevant and encountered on a daily basis, as awkward interactions could occur from different types of greeting behaviors in many situations, such as multiple individuals, varied contextual settings, and greeting styles. Hence, the stimuli of awkward greetings are more heterogeneous than natural greeting behaviors. Given the limited number of psychological studies that use naturalistic videos to examine the perception of awkwardness from daily interactions, the key perceptual signals signifying a heterogenous construct, such as awkwardness, remain unknown. Therefore, to cover a large range of awkward greeting behaviors, we included a greater number of videos featuring awkward greeting behaviors (24) than videos featuring natural greeting behaviors (10). We also removed text description in some videos (e.g., Video 22: the text caption "Trump's awkward interactions with world leaders" was removed in the RGB videos).

Procedure

Participants were presented with a randomized order of 34 raw RGB videos selected from YouTube. For each of the 34 videos (and consequently after each trial), participants were asked to categorize the greeting behavior in the video as "awkward" or "not awkward." Note that "awkward" and "not awkward" were not defined to the participant, thus allowing the participant to use their own criteria. If the participants labeled the video as awkward, they were instructed to write a brief verbal description explaining why the greeting behavior in the video appeared awkward. Participants were not given a time limit to write their descriptions. No sound was provided to the participants during video presentation.

Results

In the first step, the proportion of participants identifying greeting behaviors as awkward was reported for each video (see Fig. 2, column 5). The response proportion was used to classify each video as natural or awkward in the remaining paper. Each video was classified as natural if the mean proportion of participants categorizing it as awkward was less than .50, and a video was classified as awkward if this mean proportion was greater than .50. All 24 videos identified by the experimenters as awkward showed a mean proportion greater than .50. To assess whether people were in general agreement in judging awkwardness from human interactions, a Spearman–Brown corrected split-half reliability coefficient based participants' responses indicated high consistency ($r = .850$) across participants for detecting the presence and absense of awkward greetings in videos.

In the second step, we analyzed written descriptions for the greeting behaviors that participants categorized as awkward. Written descriptions from all the participants were merged

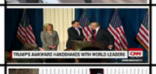
Videos classified as Awkward					Videos classified as Awkward					Videos classified as Natural				
Video ID	Frame from Video	Social Words	Motor Words	Proportion Awkward	Video ID	Frame from Video	Social Words	Motor Words	Proportion Awkward	Video ID	Frame from Video	Social Words	Motor Words	Proportion Awkward
1		26	32	1.00	13		11	21	0.87	25		24	1	0.43
2		18	11	0.97	14		16	12	0.87	26		2	2	0.40
3		29	19	0.97	15		28	11	0.87	27		5	0	0.37
4		20	22	0.93	16		56	9	0.87	28		11	12	0.33
5		38	27	0.93	17		19	16	0.83	29		5	0	0.30
6		17	29	0.93	18		27	20	0.83	30		5	4	0.23
7		25	19	0.93	19		19	9	0.80	31		0	0	0.17
8		25	24	0.93	20		18	16	0.77	32		1	1	0.13
9		41	9	0.90	21		17	13	0.77	33		1	0	0.07
10		17	21	0.90	22		8	25	0.77	34		0	0	0.03
11		26	16	0.90	23		21	13	0.77					
12		25	23	0.90	24		18	5	0.73					

Fig. 2 Examples of videos classified as awkward (left and middle panels) and natural (right panel), including key frame, corresponding frequency of social and motor word descriptions (generated from participants' written descriptions), and proportion of awkward responses (> 0.5 indicates that the greeting action in the video was perceived as being awkward, < 0.5 indicates that the perception was natural behavior)

into one file in order to identify the high-frequency words including nouns, verbs, adverbs, and adjectives. For example, sample descriptions for the awkward videos included, "*This was awkward because the man in the middle attempted to shake hands with two people at once, using his left hand for another person's right hand*" and "*The two men gestured, but it was small enough that the other person did not catch on fast enough so they were almost playing footsie with their hands.*" The text file included all participants' descriptions was entered into Textalyser (http://textalyser.net), an online software that provides a ranking of the most frequently occurring words used in a body of text. After excluding words with fewer than three characters and numerals, Textalyser returned the 200 most frequent words in the entire set of participants' written descriptions. Of those 200 words, the first two authors selected a subset of words that consisted of verbs, adverbs, and adjectives and that excluded nouns and redundant words like "awkward" or "handshake" or words in phrases like "fist bump" or "high five." From that subset, the authors selected words that were either motor related or social related. Motor-related words expressed actions predominantly related to motor behavior (e.g., "pull," "grab," "reach") or their properties (e.g., "toward"). Social-related words had a wider range and included words related to mental attributes (e.g., "try," "want," or "confuse"). We also categorized social-related words as those whose use indicates social knowledge about appropriate greetings (e.g., "kiss," "long," "far"). For example, "far" is considered a social word because there appeared to be an ideal socially acceptable distance between two people based on inherent social knowledge (too "far"). See the full list in Appendix B , as shown in the word cloud display (http://worditout.com) in Fig. 3.

The frequencies of the selected social-related and motor-related words in participants' written descriptions were then calculated for each of the 34 videos (see Fig. 2, columns 3 and 4). Note that when a social-related or motor-related word was a verb (e.g., "try"), all instances of its alternative forms (e.g., "tried," "tries," "trying") were counted as instances of that word. Videos showed different ratios between the number of motor-related words and the number of social-related words in written descriptions. For example, for Video 1, in which Donald Trump aggressively attempts to pull a somewhat stiff and reluctant Neil Gorsuch closer and closer to him, participants' descriptions included the most motor-related words (33). For Video 16, in which a man reaches for a handshake from a woman and then attempts to kiss the woman's hand, participants employed the greatest number of social-related words (56) in written descriptions.

Fig. 3 Word cloud displays both social-related and motor-related words used to freely describe awkward handshakes in Experiment 1. Font size corresponds to relative frequency

A Pearson correlation coefficient was computed to assess the relationship between the proportion of participants categorizing each video as awkward and the combined number of social-related and motor-related word descriptions for each video. We found a significant positive relationship, $r = .59$, $p = .002$, observed power = .965, suggesting that the identification of awkwardness is consistent with verbal descriptions of the videos. Additionally, while there was no significant relationship between the number of social words used in the descriptions and awkward proportion, there was a significant relationship between the number of motor words and awkward proportion ($r = .50$, $p = .012$, observed power = .863), suggesting the possible strong influence of motor cues on awkwardness judgments.

Collectively, the results of Experiment 1 revealed that people are able to judge awkwardness from observing greeting behaviors in videos, and showed general agreements in classifying awkward versus natural greetings. Furthermore, this judgment of awkwardness is likely reliant on both social and motor cues, as revealed by the prevalence of social-related and motor-related words in the semantic description of awkward greetings.

Experiment 2

Experiment 1 served as an exploratory study, demonstrating that participants are not idiosyncratic in their perceptions of awkwardness when viewing raw videos of greeting behaviors. Experiment 2 was designed to more objectively examine the awkwardness judgments by varying visual information provided in the display, and to assess the contributions from the kinematics of body movements and the visual context in perceived awkwardness of greeting behaviors.

Method

Participants

Sixty-six participants (female = 49, male = 17, mean age = 20.62 years) were recruited from the University of California, Los Angeles (UCLA) Psychology Subject pool. All but two participants had lived in the United States for at least five years and were fluent English speakers. The sample size was determined in accordance with a previous study on individual differences in biological motion perception (van Boxtel, Peng, Su, & Lu, 2016). Participants provided informed consent, as approved by the UCLA Institutional Review Board (#16-001879) and were given course credit for their participation.

Stimuli and apparatus

The identical 34 videos from Experiment 1 were also used to generate the stimuli in Experiment 2. The visual stimuli were presented on the center of the screen at the size of 250 × 250 pixels using MATLAB (R2017a; The MathWorks, Natick, MA). The greeting behavior stimuli were manipulated by one of the three display types: patch, body, or skeleton display (see Fig. 1 for sample image frames).

Patch display The superpixel algorithm segmented raw images into patches by grouping pixels with similar brightness, color, and texture in the local regions of images. The output of this algorithm displayed patches of the locally consistent pixels with similar colors or gray levels (Ren & Malik, 2003). As shown in the second column of Fig. 1, MATLAB's "superpixel" function was used to process the 34 raw videos to generate the patch display as a blurred visual scene. Since the patch display included blurred backgrounds and other people/objects in the display, the stimuli in the patch display provided the most contextual information in the visual scene.

Body display We used a deep learning model to segment human bodies in the video. The model, RefineNet, exploits visual features at multiple levels of abstraction for high-resolution semantic segmentation. The model classified the pixels of each video frame as belonging to a human form or not and further classified pixels corresponding to humans into the following body regions: head, torso, upper arm, lower arm, upper leg, and lower leg (Lin, Milan, Shen, & Reid, 2017). RefineNet processed the 34 raw videos to present

colored human figures against a black background, as shown in the third column of Fig. 1.

Skeleton display We used a multiperson pose estimation algorithm (Cao, Simon, Wei, & Sheikh, 2017) to estimate the location of key body joints in videos. This deep learning model detects body parts and is robust against body occlusion and viewpoints. Based on the inferred joint locations, skeleton figures were generated using BioMotion toolbox (van Boxtel & Lu, 2013) to extract the kinematic movement from the raw 34 videos, as shown in the fourth column of Fig. 1. The white skeleton was displayed against a black background. At the time of conducting the experiment (July 2017), the pose estimation algorithm did not include options to infer the hands of the actors, and only provided estimation of joint coordinates up until the wrist for arms. Therefore, the skeleton actors did not directly touch in the stimuli. Participants were informed of this display feature in order to minimize the surprise that the two hand-shaking actors did not touch the other person's skeleton. Additionally, the model occasionally failed to extract lower body parts, largely due to the similar color of pants as the background, or the missing parts (such as lower legs) occasionally occluded by objects (such as a table) in the YouTube videos. To correct for this, and maintain consistency across videos, a gray rectangular occluder was displayed at the bottom of the screen, which covered missing body parts.

Procedure

Participants were randomly assigned to view one of three display types (i.e., patch, body, skeleton). Participants first viewed a sample video (which was manipulated by the assigned display type and was not included in the experimental test trials), in order to gain familiarity with the display type. Following exposure to the sample video, participants began the experiment. In each trial, participants were asked to rate the degree of awkwardness of the greeting behavior in the video stimulus on a 6-point scale from 1 (*surely natural*), 2 (*probably natural*), 3 (*guess natural*), 4 (*guess awkward*) and 5 (*probably awkward*) to 6 (*surely awkward*). The subjects were not imposed with a time limit during the rating period and no sound was provided to the participants from the videos. The experiment consisted of 34 trials with randomized order and lasted around 30 minutes.

Results

Mean human ratings for the three display types in Experiment 2 were significantly correlated with the proportion of awkward responses for the raw video recordings in Experiment 1 (r = .92 for patch, observed power = .999, r = .83 for body, observed power = .999, and r = .84, observed power = .999 for skeleton, ps < .001), suggesting that participants were in general agreement about categorizing awkward or natural interactions across all displays.

Next, we conducted a mixed ANOVA with one within-subjects factor, activity normality (awkward vs. natural greetings) and one between-subjects factor, display type (patch vs. body vs. skeleton). The activity normality of videos was determined by the proportion of responses in Experiment 1 that classified the videos as awkward (proportion > 0.5) or natural (proportion < 0.5). The average ratings for awkward videos and natural videos were used as dependent variables in the ANOVA analysis. As shown in Fig. 4, a main effect of activity normality was revealed to show higher ratings for awkward videos than for neutral videos, $F(1, 63) = 539.100$, $p < .001$, $\eta_p^2 = .895$. We also found a significant two-way interaction effect between activity normality and display types, $F(2, 63) = 11.096$, $p < .001$, $\eta_p^2 = .260$. Specifically, the impact of display type on ratings was not found for natural videos (patch vs. body; patch vs. skeleton, ps > .05). However, for awkward videos,

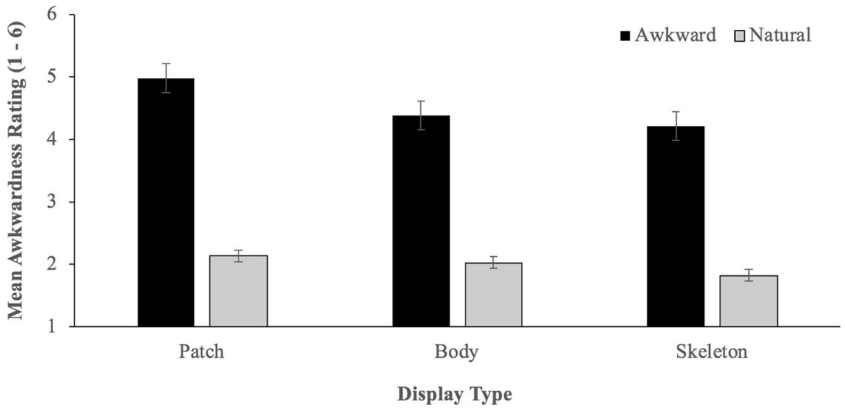


Fig. 4 Mean awkwardness ratings for awkward and natural videos as a function of three types of displays. Error bars indicate standard error of the means

participants yielded greater awkwardness ratings for the patch display than for the other two displays (patch vs. body, $p = .003$; patch vs. skeleton, $p < .001$, with Bonferroni correction). Since both the skeleton and body displays remove a majority of the contextual information (i.e., actor identity and scene depiction), the lower awkwardness ratings in these two displays reveal the strong influence of contextual information about actor characteristics and scene background in awkwardness judgments. This impact of display type on ratings for awkward videos is also consistent with the observations that awkwardness perceived in some videos differs depending on the display type. For example, the famous video of President Donald Trump shaking hands with Neil Gorsuch(Video 1) was ranked highly awkward in the body display (rank #5; ranging from 1, *most awkward*, to 34, *least awkward*). But in the skeleton and patch display, people gave slightly lower awkwardness ratings, although still considered awkward (#12 for skeleton, #8 for patch). Another example is Video 7. This video involves a scene where a person is intentionally avoiding a second person's high five, fist bump, and hug. Here, the video is consistently rated as awkward in the displays with increased contextual information (as rank #4 in the patch display and rank #5 in the body display). However, the skeleton display of the video was no longer rated as high on awkwardness (rank #13). This rating difference likely results from the minimum contextual information in the skeleton display. This result is consistent with participants' written descriptions when viewing Video 7 with 65% more social words than motor words, suggesting social context may play an important role in judging awkwardness for this video.

Since our experiment consisted of greeting behaviors with different durations, it is possible that people use a simple heuristic relying on greeting durations to judge the awkwardness. To address this possibility, we conducted a multiple regression analysis to assess whether video duration significantly accounted for variability in participants' awkwardness ratings. We used participants' mean awkwardness ratings for the patch, body, and skeleton displays as the outcome variable, and used two predictor variables: the proportion of awkward responses for each raw RGB video (measured in Experiment 1) and video duration (measured in seconds). For all three display types, awkwardness ratings were predicted by the proportion of participants classifying each raw RGB video, but not by video durations. Specifically, for the patch display, a regression model was significant, $F(2, 31) = 67.929$, $p < .001$, with an $R^2 = .814$. However, there was no relationship between video duration and mean awkwardness ratings for the patch display ($\beta = .094$, *ns*), while the proportion of participants classifying each video as awkward served as a significant predictor ($\beta = .861$, $p < .001$). For the body displays, a significant regression equation was found, $F(2, 31) = 27.514$, $p < .00$ with an $R^2 = .739$. Similarly, the proportion of participants classifying the video as awkward was a significant predictor ($\beta = .796$, $p < .001$), while there was no linear relationship between video duration and mean awkwardness ratings for the body display ($\beta = .136$, *ns*). For the skeleton display, a significant regression model was also found, $F(2, 31) = 37.420$, $p < .001$, with $R^2 = .707$. Additionally, the proportion of participants classifying the video as awkward served as a significant predictor ($\beta = .773$, $p < .001$), while there was no significant relationship between video duration and mean awkwardness ratings for the body display ($\beta = .146$, $p < .001$), consistent with model results from the patch and body display. Therefore, the varied video duration did not appear to influence participant's awkwardness judgments.

To better understand a psychological space underlying the awkwardness judgments, we conducted a multidimensional scaling (MDS) analysis (Kruskal & Wish, 1978) to explore what psychological dimensions play key roles in determining participants' ratings for awkwardness. We first included all 34 videos in the MDS analysis; however, the MDS results appeared to cluster all the awkward videos in similar locations to separate from the natural videos, which are not informative for visualizing the basic features sensitive to the different degrees of awkward behavior. Hence, in the final MDS analysis, we only included ratings for the 24 awkward videos identified in Experiment 1. We computed the Euclidean distance between any pairs of ratings for awkward videos to generate the distance matrix for the 24 videos for each display type. Smaller distances reflected that the pair of actions were judged with similar awkwardness ratings across subjects. The 24 × 24 distance matrix was the input for the nonmetric MDS to project the pairwise distances of awkwardness ratings to a two-dimensional space.

As shown in Fig. 5, the resultant space of MDS analysis was a two-dimensional psychological space with $r^2 = .94, .92, .83$; and stress = .07, .10, .17 for the displays of patch, body, and skeleton, respectively. Using a two-dimensional space was adequate as adding more dimensions just provided marginal improvements in stress as shown in Fig. 6. Across all display types, we found a significant correlation between the horizontal coordinates of videos and the number of social words present from descriptions of the videos in Experiment 1 (for patch display, $r = -.488$, $p = .015$, observed power = .686; body display, $r = -.562$, $p = .004$, observed power = .514; skeleton display, $r = -.611$, $p = .002$, observed power = .903). These correlation results were consistent with individual observation of clusters in Fig. 5, which revealed that actions exhibiting a higher degree of social incoordination (with more social word descriptors) were consistently located at the left end of the resultant psychological space in the MDS result plot. For example, Video 16 was consistently located on the left side

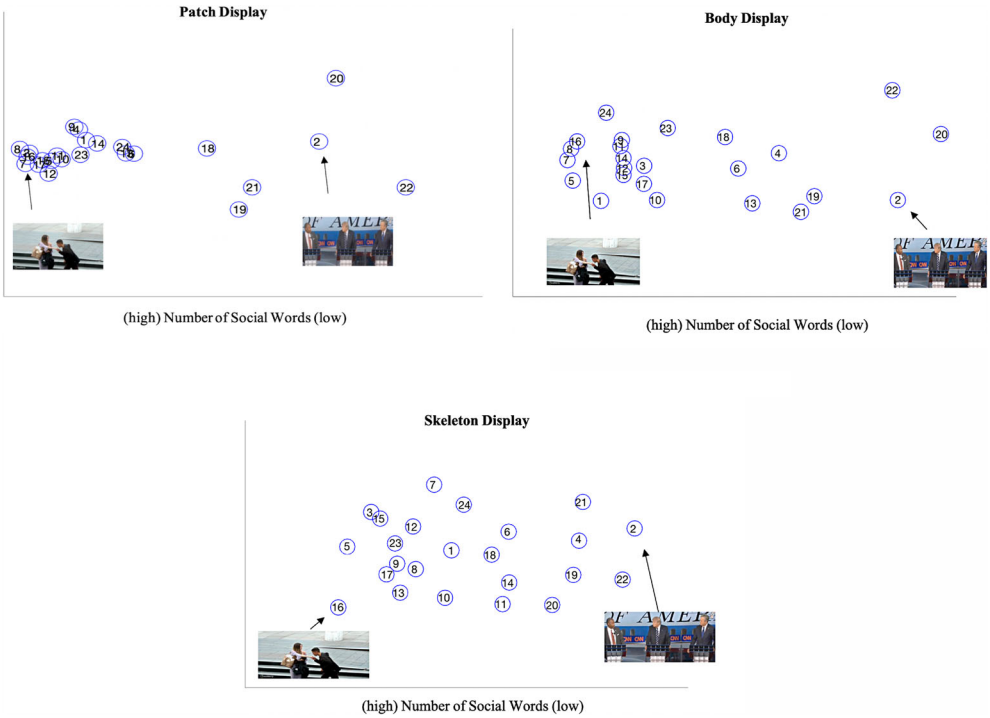


Fig. 5 Results of the psychological space from MDS analysis for patch display (top left panel), body display (top right panel), and skeleton display (bottom panel). Video 16 (boy kisses girl's hand) and Video 2 (Donald Trump's aggressive handshake/grab) are representative of videos consistently appearing in the similar horizontal location across display types, with Video 16 having more social descriptors and Video 2 having less social descriptors

of the MDS result plot, wherein a man attempted to kiss (in lieu of handshaking) a young woman who avoided the body contact. This action involves a high degree of social incoordination due to the lack of engagement of the female actor, and a general violation of social heuristics (kissing instead of handshaking). On the other hand, awkward videos with a lower degree of social incoordination (determined by their lower number of social descriptors from Experiment 1) were clustered on the right side. For example, Video 2, showing President Trump catching and vigorously shaking Ben Carson's hand at a presidential debate, was rated as highly awkward for the raw video. Even after removing the identity information, the handshaking videos in all the three displays remained on the right side of the MDS space, due to the low degree of social incoordination in their interaction. Hence, the horizontal dimension in the psychological space of awkwardness judgment reveals the perceived degree of social incoordination, an overall impression of how well the two actors coordinate their social interaction in the greeting behavior.

The first dimension (horizontal) accounted for most of the variability in the awkwardness judgments across the videos**:** 86%, 79%, and 50% of the variance for the displays of patch, body, and skeleton, respectively. In contrast, the vertical dimension accounted for less variability, as 44%, 21%, and 16% of the variance for the three displays. The interpretation of the vertical dimension is not as clear as the horizontal dimension. Because the number of social words for each video from Experiment 1 correlated with the horizontal coordinates, we explored whether the vertical coordinates correlated with the number of motor words for each video from Experiment 1, but did not find the relationship across all display types. However, we noticed a possible relation with the touching duration for the skeleton display. For each video, we estimated whether "touching," defined by the distance between the two wrist points, was less than the average lower arm length. We found that the vertical locations of videos in the psychological space showed a marginal correlation with touching duration in the greeting behaviors in the skeleton display, r = .390, p = .060, observed power = .502. We conjecture that when only body kinematics are available in the skeleton input, physical contact may serve as an important cue for signaling motor coordination between actors, likely related to internal knowledge regarding the appropriate duration of touching in the present greeting behaviors. However, when more contextual information is available in greeting behaviors in the patch and body displays, other contextual cues, aside from touching duration, may jointly affect awkwardness judgments.

General discussion

Humans encounter and experience awkward social interactions on a daily basis, yet previous research has neither

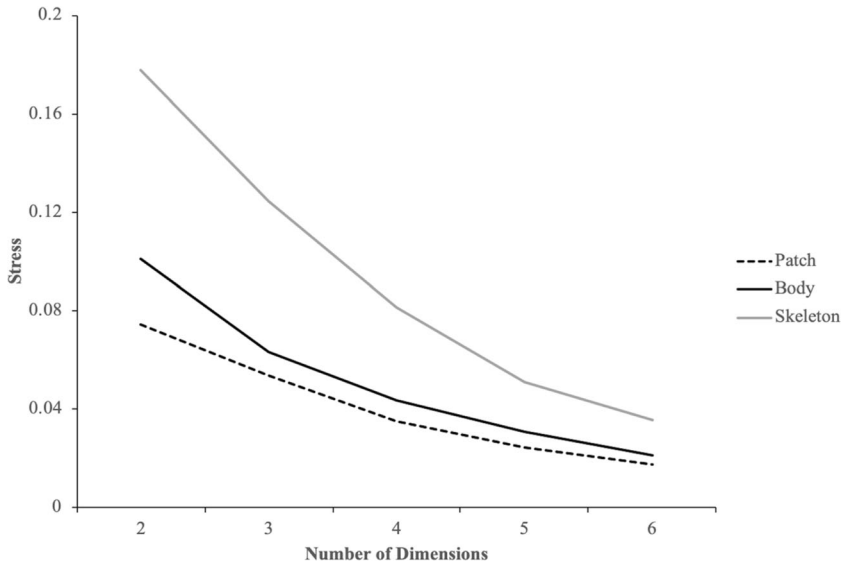


Fig. 6 Stress plot of MDS representation for each of the three display types. A two-dimensional space was selected as the best choice because there was only a marginal improvement in stress with increased dimensionality

sufficiently explored this complex social construct nor investigated the contribution of kinematics to these judgments. Largely absent is an interdisciplinary "person construal" approach (Freeman & Ambady, 2011), relating lower level perceptual mechanisms (e.g., kinematics) with attributions of higher level social judgments of awkwardness. Therefore, in two experiments we examined whether awkwardness is inherently subjective, or whether there exists a more objective, streamlined criterion that humans use to reliably categorize awkwardness. Experiment 1 served as a preliminary study examining whether awkwardness can be reliably judged from greeting behaviors, and how this ability related to the presence of social and motor cues. Using the written descriptors from Experiment 1, we rank-ordered the videos based on the proportion of participants classifying each interaction as awkward. In Experiment 2, we manipulated the amount of visual information while holding body kinematics constant. Together, the present experiments revealed that participants were systematically able to judge awkward behaviors across all three display types, underscoring the potential importance of human kinematics to social interpretations and its key role in signaling awkward behavior. Importantly, this ability appeared to account for the presence of contextual information (body morphology, actor identity, and scene depiction), as revealed by significantly greater awkwardness ratings for the patch display (with the highest degree of contextual information) than the body or skeleton displays.

To compare these results with individual cases, we examined particular videos judged with higher awkwardness ratings. We found that the video of President Donald Trump shaking hands with Neil Gorsuch (Video 1) was rated consistently higher (i.e., more awkward) than many of the other videos, even when visual cues to identity were removed, or reduced (as in the skeleton and body displays), suggesting that the perception of President Trump's awkward handshaking may primarily be attributed to his motoric "awkward" behavior (barring external influences of contextual differences). While visual identity generally appears to play a predominant role in social judgments (e.g., knowledge of President Trump) as well as in action recognition (e.g., Ferstl, Bulthoff, & de la Rosa, 2016), even observing the kinematics of human movements is a viable tool to make high-level social judgments of interactions. In fact, participants' written descriptions on the raw video corroborated this finding by including a doubled number of motor descriptions than social descriptions in their awkwardness descriptions.

Previous research has found that certain human motor cues play an important role in the detection of threatening actions (van Boxtel & Lu, 2011, 2012), perception of social interactions (Thurman & Lu, 2014), emotion perception from actions (Roether, Omlor, Christensen, & Giese, 2009), and action discrimination (van Boxtel & Lu, 2015). What specific cues affect the perception of awkwardness in social interactions? To probe the

underlying psychological dimensions of awkwardness judgments in greeting behaviors, MDS analysis revealed two important candidates: social incoordination and touching duration (overall length of handshake/greeting). Social incoordination accounted for most of the variability in judging awkwardness. Here, we define social (in)coordination as the degree to which individual actions are (not) in accordance with greeting behavior norms (e.g., shaking hands instead of kissing) and also with the physical setting in which the interaction takes place. Individual cases, such as President Trump catching and shaking Ben Carson's hand, demonstrate how (although ranked awkward) the interaction does not violate social coordination within the presidential debate setting (clustered on the right, or congruent, side). However, in cases of the first-time meeting with another individual in a public setting, a violation of social appropriateness in American culture is likely to occur when a stranger attempts to grab and kiss another individual's hand (as seen in Video 16 clustered on the opposite side in psychological space).

Through visual inspection, we also found that actions with an obvious motor incoordination (e.g., missed catch) also tended to cluster on the left side of the MDS space. Videos in this cluster consisted of strong motor incoordination (e.g., Video 5 featuring a dyadic interaction consisting of a series of missed fist bumps and handshakes). Meanwhile, the opposite (right side) of the horizontal dimension showed the cluster of actions with good motor coordination (e.g., Video 22 showing Donald Trump shaking Mitt Romney's hand) in the greeting behavior, suggesting that the degree to which the two actors coordinate their movements in the display can signal a key underlying dimension of whether an interaction is awkward or not. Given the relationship between the horizontal dimension and the number of social words in the descriptions, we conjecture that motor coordination likely factors into participants' social judgments since motor coordination does not generally occur in isolated situations.

Touching duration may serve as a secondary cue for signaling awkward greeting behaviors. While cautious in our interpretation of the marginally significant relationship (Pritschet, Powell, & Horne, 2016), this result is still consistent with previous findings that people are sensitive to temporal relation between actors (Burling & Lu, 2018; de la Rosa, et al., 2014; Sebanz & Knoblich, 2009) and different motion cues in actions (Peng, Thurman, & Lu, 2017; Thurman & Lu, 2016). Furthermore, interpersonal touch also serves as a nonverbal social cue, incorporating important social information, such as emotional attributes and bonding (Gallace & Spence, 2010). Our results suggest that the correlations observed in the patch (most contextual information) and skeleton (least contextual information), but not in the body display (medium degree of contextual information) may be due the importance of touching duration as a social cue to awkwardness when incorporating key contextual information, such as the setting in which an interaction occurs (as in the patch display) and the importance of touching duration as a motor cue to awkwardness when relying predominantly on human kinematic information (as in the skeleton display). Further characterizing this relationship as it pertains to cultural differences, or remains inherent to American society, is an interesting area of future exploration.

As an final point, while we aimed to separately examined the contribution of contextual information, consisting of rich social cues (e.g., scenery, attire), and human kinematic information (consisting of rich motor cues), as key signals underlying the evaluation of awkward interactions, the MDS results reveal their inextricable link. Specifically, the similar clustering of videos with both social and motor incoordination, as well as the key motor signals in touching duration (also related to social heuristics), prompt the following question: To what extent is motor coordination distinct from social coordination? Literature on interpersonal social interactions has shown that the temporal and motion congruency between two agents underlie human perceptions of social traits and/or animacy to the interaction (Thurman & Lu, 2014). Our present results similarly converge, suggesting that social coordination is likely affected by motor coordination in awkwardness judgments. Although objectively examining the extent of this relationship is outside the scope of our present paper, these results point to an important area of investigation that can even extend to wide-ranging, more ecologically valid domains, including human–robot interaction.

We conclude that the perception of awkwardness in greeting interactions is based on general principles that significantly rely on motor cues, with the additional detection of failed social coordination for body movements that provide a key signal that a greeting has gone awry. Importantly, judging awkwardness does not appear to be entirely idiosyncratic—individuals appear to predominantly rely on a general set of heuristics rooted in human kinematics that is dynamically coupled with contextual information.

Acknowledgements This study was supported by the National Science Foundation (BCS-1655300). We would like to thank Tianmin Shu with help with skeleton extraction, and Tabitha Safari, Justin Azarian, Sophia Baia, Devin Bennett, and Jaime Wu for help with data collection.

The data and materials for all experiments are available online on Mendeley (DOI: https://doi.org/10.17632/mbz4n7n5m8.1). Neither experiment in our study was preregistered.

Appendix A

Table 1 Links to video stimuli

Clip name	Link	Start	Duration (s)
Clip1	https://youtu.be/T84se4fc4KU?t=30s	0:31	6.72
Clip2	https://youtu.be/_npyQonU-V0?t=1m39s (N/A)	4:72	3.80
Clip3	https://www.youtube.com/watch?v=i6i1PpEtM_4	1:06	10.04
Clip4	https://youtu.be/6yWCmMQy1Qk?t=1m45s	1:44	3.80
Clip5	https://www.youtube.com/watch?v=c5aN5zOEpM8	0:05	26.96
Clip6	https://youtu.be/T84se4fc4KU?t=28s	0:27	3.08
Clip7	https://www.youtube.com/watch?v=c5aN5zOEpM8	1:09	12.04
Clip8	https://www.youtube.com/watch?v=c5aN5zOEpM8	1:21	7.64
Clip9	https://youtu.be/_npyQonU-V0?t=5m03s (N/A)	5:03	7.92
Clip10	https://youtu.be/_npyQonU-V0?t=9m6s (N/A)	9:06	7.16
Clip11	https://www.youtube.com/watch?v=c5aN5zOEpM8	1:01	6.36
Clip12	https://www.youtube.com/watch?v=i6i1PpEtM_4	2:43	10.40
Clip13	https://youtu.be/wT9Prne9wF0 (N/A)	0:00	6.40
Clip14	https://youtu.be/_npyQonU-V0?t=31s (N/A)	0:31	8.04
Clip15	https://www.youtube.com/watch?v=i6i1PpEtM_4	1:42	8.04
Clip16	https://www.youtube.com/watch?v=i6i1PpEtM_4	2:35	3.68
Clip17	https://www.youtube.com/watch?v=i6i1PpEtM_4	2:24	7.72
Clip18	https://www.youtube.com/watch?v=i6i1PpEtM_4	2:52	4.40
Clip19	https://youtu.be/iPDM0msZwQk?t=8s	0:08	10.24
Clip20	https://youtu.be/_npyQonU-V0?t=1m8s (N/A)	1:08	4.80
Clip21	https://youtu.be/_npyQonU-V0?t=3m50s (N/A)	03:50	5.24
Clip22	https://youtu.be/T84se4fc4KU?t=24s	0:23	3.68
Clip23	https://www.youtube.com/watch?v=i6i1PpEtM_4	0:45	2.76
Clip24	https://www.youtube.com/watch?v=c5aN5zOEpM8	0:51	3.72
Clip25	https://www.youtube.com/watch?v=JPhIPT9yOu8	0:42	8.00
Clip26	https://youtu.be/WmuybcgSjkI (N/A)	3:24	3.20
Clip27	https://youtu.be/_npyQonU-V0?t=47s (N/A)	0:47	3.60
Clip28	https://youtu.be/09ZtoTthtys?t=4m50s	4:50	5.56
Clip29	https://youtu.be/HVK-xbdddhA?t=3m8s	3:07	6.32
Clip30	https://youtu.be/09ZtoTthtys?t=9m13s	9:14	3.20
Clip31	https://youtu.be/V-mR66UW8NI?t=105	1:45	2.08
Clip32	https://youtu.be/09ZtoTthtys?t=4m57s	4:57	3.00
Clip33	https://www.youtube.com/watch?v=muN0yh_STkM	0:05	2.64
Clip34	https://youtu.be/V-mR66UW8NI?t=30s	0:30	4.04

[+] As of publication, please note that some links no longer exist since they were collected in 2017. Due to copyright issues, we cannot upload the full videos publicly. We have indicated with N/A next to the links that do not exists. Please do not hesitate to contact the authors if interested in viewing the original video

Appendix B

Table 2 List of motor and social words classification in Experiment 1

Word	Category	Frequency
try	social	111
want	social	84
long	social	55
attempt	social	39
know	social	25
uncomfortable	social	23
confused	social	20
kiss	social	20
think	social	18
give	social	16
initiating	social	10
expecting	social	9
pull	motor	45
reach	motor	39
hold	motor	38
gesture	motor	38
grab	motor	36
change	motor	35
move	motor	23
touch	motor	20
pull	motor	20
times	motor	19
far	motor	19
towards	motor	19
turned	motor	17
timing	motor	11
respond	motor	11
playing	motor	11
extended	motor	10
continued	motor	9
switched	motor	9

References

Alter, A. L., Stern, C., Granot, Y., & Balcetis, E. (2016). The "bad is black" effect: Why people believe evildoers have darker skin than do-gooders. *Personality and Social Psychology Bulletin*, *42*(12), 1653–1665. doi:https://doi.org/10.1177/0146167216669123

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*(4), 599–614. doi:https://doi.org/10.1037/0022-3514.86.4.599

Burling, J. M., & Lu, H. (2018). Categorizing coordination from the perception of joint actions. *Attention, Perception, & Psychophysics, 80*(1), 7–13. doi:https://doi.org/10.3758/s13414-017-1450-2

Burling, J. M., Kadambi, A., Safari, T., & Lu, H. (2018). The impact of autistic traits on self-recognition of body movements. *Frontiers in Psychology, 9,* 2687

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *The IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7291–7299). doi:https://doi.org/10.1109/CVPR.2017.143

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the psychonomic society, 9,*(5), 353–356

de la Rosa, S., Choudhery, R. N., Curio, C., Ullman, S., Assif, L., & Bülthoff, H. H. (2014). Visual categorization of social interactions. *Visual Cognition, 22*(9), 1233–1271. doi:https://doi.org/10.1080/13506285.2014.991368

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247–279. doi:https://doi.org/10.1037/a002232

Freeman, J. B., Ma, Y., Han, S., & Ambady, N. (2013). Influences of culture and visual context on real-time social categorization. *Journal of Experimental Social Psychology, 49,* 206–210. doi:https://doi.org/10.1016/j.jesp.2012.10.015

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLOS ONE*, *6*, e25107. doi:https://doi.org/10.1371/journal.pone.0025107

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: A naturalistic measure of social understanding in autism. *Journal of Autism and Developmental Disorders*, *30*(3), 225–236. doi:https://doi.org/10.1023/a:1005544518785

Hohmann, T., Troje, N. F., Olmos, A., & Munzert, J. (2011). The influence of motor expertise and motor experience on action and actor recognition. *Journal of Cognitive Psychology, 23*(4), 403–415. doi:https://doi.org/10.1080/20445911.2011.525504

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research, 38*, 379–393. doi:https://doi.org/10.1007/BF00309043

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving sex directly and indirectly: Meaning in motion and morphology. *Psychological Science, 16*(11), 890–897.

Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling (SAGE University Paper Series on Quantitative Applications in the Social Sciences, No. 07–011). Beverly Hills, CA: SAGE. doi:https://doi.org/10.4135/9781412985130

Lin, G., Milan, A., Shen, C., & Reid, I. (2017). RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1925–1934). doi:https://doi.org/10.1109/CVPR.2017.549

Peng, Y., Thurman S., & Lu, H. (2017). Causal action: a fundamental constraint on perception and inference with body movements. *Psychological Science, 28*(6), 789–807.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82*(2), B51–B61. doi: https://doi.org/10.1016/s0010-0277(01)00147-0

Ren, X., & Malik, J. (2003). Learning a classification model for segmentation. *Proceedings of the Ninth International Conference on Computer Vision* (pp. 10–17). doi:https://doi.org/10.1109/ICCV.2003.1238308

Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision, 9*(6), 15–15. doi:https://doi.org/10.1167/9.6.15

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception:

Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General, 112*(4), 585–615. doi:https://doi.org/10.1037/0096-3445.112.4.585

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50-ms. *Journal of Experimental Social Psychology, 44,* 1100–1105. doi:https://doi.org/10.1016/j.jesp.2007.12.001

Rule, N. O., & Ambady, N. (2010). Democrats and Republicans can be differentiated from their faces. *PLOS ONE, 5,* e8733. doi:https://doi.org/10.1371/journal.pone.0008733

Rule, N. O., Ambady, N., Adams, R. B., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology, 95,* 1019–1028. doi: https://doi.org/10.1037/a0013194

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human Movement Science*, *30*(5), 834–845. doi:https://doi.org/10.1016/j.humov.2010.05.014

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science, 1*(2), 353–367. doi: https://doi.org/10.1111/j.1756-8765.2009.01024.x

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*(9), 656–666. doi:https://doi.org/10.1037/0022-3514.35.9.656

Thurman, S., & Lu, H. (2014). Perception of social interactions for spatially scrambled biological motion. *PLOS ONE, 9*(11), 1–12. doi: https://doi.org/10.1371/journal.pone.0112539

Thurman, S., & Lu, H. (2016). Revisiting the importance of common body motion in human action perception. *Attention, Perception, & Psychophysics, 78*(1), 30–36. doi:https://doi.org/10.3758/s13414-015-1031-1

Todorov, A., Said, C. P., & Verosky, S. C. (2011). Personality impressions from facial appearance. In A. Calder, J. V. Haxby, M. Johnson, & G. Rhodes (Eds.), Handbook of face perception (pp. 631–652). Oxford, England: Oxford University Press. doi:https://doi.org/10.1093/oxfordhb/9780199559053.013.0032

van Boxtel, J., & Lu, H. (2011). Visual search by action category. *Journal of Vision. 11*(7), 1–14. doi:https://doi.org/10.1167/11.7.19

van Boxtel, J., & Lu, H. (2012). Signature movements lead to efficient search for threatening actions. *PLOS ONE, 7*(5), e37085, 1–6. doi: https://doi.org/10.1371/journal.pone.0037085

van Boxtel, J., & Lu, H. (2013). A biological motion toolbox for reading, displaying and manipulating motion capture data in research settings. *Journal of Vision, 13*(12), 7, 1–16. doi:https://doi.org/10.1167/13.12.7

van Boxtel, J., & Lu, H. (2015). Joints and their relations as critical features in action discrimination: Evidence from a classification image method. *Journal of Vision, 15*(1), 20, 1–17. doi:https://doi.org/10.1167/15.1.20