Adaptation aftereffects reveal representations for encoding of contingent social actions

Leonid A. Fedorov[a,b,1], Dong-Seon Chang[b,c], Martin A. Giese[a,b], Heinrich H. Bülthoff[c,2], and Stephan de la Rosa[c,1,2]

[a]Section for Computational Sensomotorics, Department of Cognitive Neurology, Centre for Integrative Neuroscience & Hertie Institute for Clinical Brain Research, University Clinic Tübingen, 72076 Tübingen, Germany; [b]International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, 72076 Tübingen, Germany; and [c]Department of Human Perception, Cognition, and Action, Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

A hallmark of human social behavior is the effortless ability to relate one's own actions to that of the interaction partner, e.g., when stretching out one's arms to catch a tripping child. What are the behavioral properties of the neural substrates that support this indispensable human skill? Here we examined the processes underlying the ability to relate actions to each other, namely the recognition of spatiotemporal contingencies between actions (e.g., a "giving" that is followed by a "taking"). We used a behavioral adaptation paradigm to examine the response properties of perceptual mechanisms at a behavioral level. In contrast to the common view that action-sensitive units are primarily selective for one action (i.e., primary action, e.g., 'throwing"), we demonstrate that these processes also exhibit sensitivity to a matching contingent action (e.g., "catching"). Control experiments demonstrate that the sensitivity of action recognition processes to contingent actions cannot be explained by lower-level visual features or amodal semantic adaptation. Moreover, we show that action recognition processes are sensitive only to contingent actions, but not to noncontingent actions, demonstrating their selective sensitivity to contingent actions. Our findings show the selective coding mechanism for action contingencies by action-sensitive processes and demonstrate how the representations of individual actions in social interactions can be linked in a unified representation.

perception | vision | social interaction | action recognition | adaptation

Adaptation effects have been demonstrated to be a powerful tool in the examination of response properties of neural processes at the behavioral level. Specifically, behavioral adaptation can selectively target the neural mechanisms and tuning characteristics of perceptual processes across the cortical hierarchy (1–4). The inference about the neural mechanisms from the behavioral adaptation is grounded in electrophysiological studies showing how repetitive sensory stimulation results in a transient response decrease in the neuronal populations involved in the processing of the stimulus (5–12). As a behavioral consequence of this neural mechanism, the perception of a subsequently presented ambiguous test stimulus is altered for a short period. Therefore, it is often fruitful to first find and investigate a behavioral adaptation effect before examining in detail the underlying neural mechanisms.

Results

Recently, we developed an experimental paradigm with which we can probe the ability to discriminate different actions using adaptation (action discrimination adaptation). We have previously shown that adaptation to an action (e.g., "throwing") transiently changes the categorical visual percept of a subsequently presented ambiguous action that contains visual elements of two actions [e.g., a morph between a throwing and a "giving" (Movie S1) or a "catching–taking" pair (Movie S2)] (13, 14). For example, adapting to a throwing action transiently causes participants to report an ambiguous appearing test action (e.g., a "throwing–giving" morph) to look more like giving and vice versa.

In the current study, we used this paradigm to examine the underlying representation of spatiotemporally proximate actions (contingent actions). This representation might be the basis of the human ability to relate one action in a social interaction, e.g., throwing, to another, e.g., giving. Specifically, we were interested in whether action representations are sensitive to both their primary action and to the related contingent action. To this end, we used two actions, namely throwing and giving, to examine the sensitivity of these action representations to the contingent actions of catching and taking. In a standard version of this adaptation paradigm, adapting to a throwing action causes participants to perceive an ambiguous action morph between throwing and giving more likely as giving. We reasoned that if the representation of a throwing action is also sensitive to the contingent action (i.e., catching), then the catching adaptor should also induce an adaptation aftereffect when observing a throwing action. Specifically, adaptation to catching should lead participants to recognize a throwing–giving test morph action more likely as a giving. That is, if action representations are also sensitive to the contingent action, then a catching adaptor will cause a similar (yet likely smaller) adaptation effect to a throwing adaptor when testing with a throwing–giving morph action (see Fig. 1B for a common adaptation experiment scheme). We refer to these adaptation effects induced by the contingent action hereafter as "cross-adaptation effects" (see Fig. 1A for cross- and normal adaptation conditions). The presence of significant cross-adaptation effects provides direct evidence for action representations being sensitive to their contingent action.

We used two social interactions, throwing–catching and giving–taking (Fig. 2A and Movies S1 and S2), to create ambiguous test stimuli. Specifically, we morphed (Methods) between

Significance

Why is it so easy for humans to interact with each other? In social interactions, humans coordinate their actions with each other nonverbally. For example, dance partners need to relate their actions to each other to coordinate their movements. The underlying neurocognitive mechanisms supporting this ability are surprisingly poorly understood. We show that human brain processes are sensitive to pairs of matching actions that make up a social interaction. These findings provide insights into the perceptual architecture that helps humans to relate actions to each other. This capability is essential for social interactions, and its understanding will aid future development of therapies to treat social cognitive disorders.

[1]L.A.F. and S.d.l.R. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: heinrich.buelthoff@tuebingen.mpg.de or delarosa@tuebingen.mpg.de.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

NEUROSCIENCE

A

Giving   Taking

Giving-throwing
morph axis

cross

normal

Throwing   Catching

Taking-catching
morph axis

B

4x

adaptator   blank   blank   test
ISI   ISI   (ambiguous morph)

1.1s   0.25s   0.4s   0.4s

time
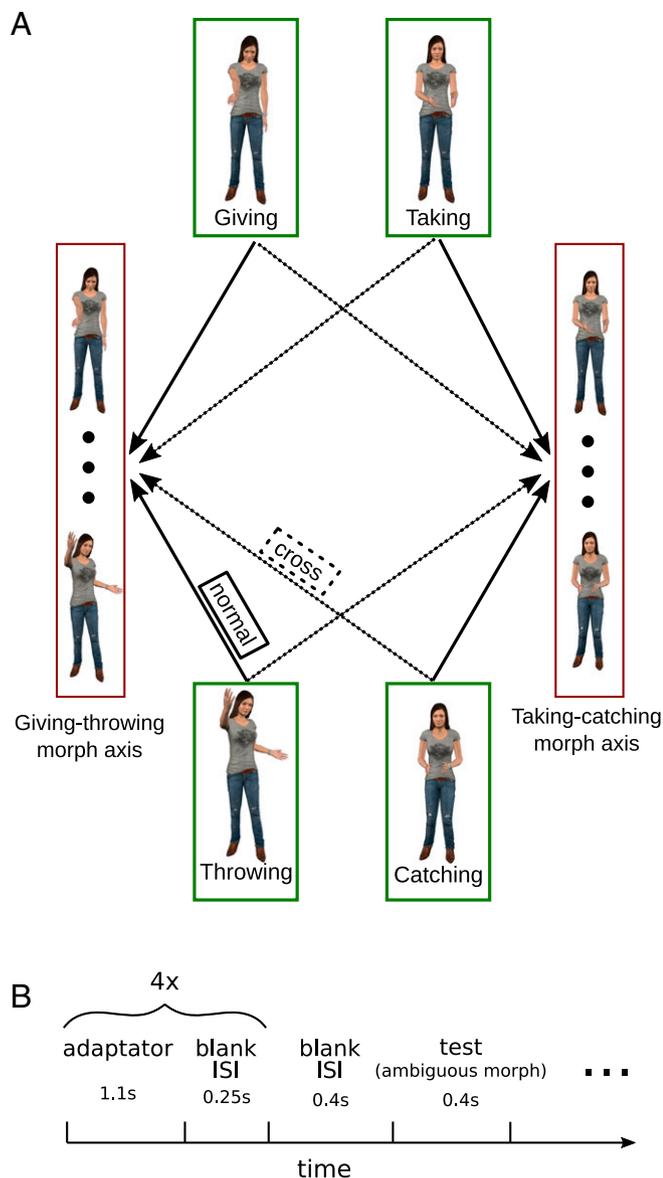
Fig. 1. (*A*) Logic of experiment 1. The experiment consists of four unmorphed adaptor actions (green squares), and two morph axes from which test stimuli were sampled. Each axis morphed between two actions: throwing and giving (*Left* red square); and catching and taking (*Right* red square). Experiment 1 combined adaptors and morph axes in different ways. Under normal adaptation conditions, the test stimuli were generated by morphing between the adapted actions (solid arrows). In the cross-adaptation condition, the test stimuli were generated from the other two adaptor actions that were not used during adaptation (dotted lines). (*B*) Schematic outline of a single experimental trial: four adaptor repetitions preceded the ambiguous test stimulus presentation. Depending on the condition, the participants' task was to indicate, for example, whether the test stimulus looked more like a give or a throw action (see *Methods* for details).

throwing and giving (i.e., along the throwing–giving morph axis) and between catching and taking (i.e., along the catching–taking morph axis). In experiment 1, we asked participants (*n* = 25) to categorize these test morphs after repeated presentation of the four different adaptor actions (catching, taking, throwing, giving). Under normal adaptation conditions, the test and adaptor stimuli were taken from the same morph axis (e.g., using giving as the adaptor stimulus and throwing–giving morph as the test stimulus). Under the cross-adaptation conditions, the adaptor and test were taken from action-contingent morph axes (e.g.,

using taking as the adaptor and throwing–giving as the test). We hypothesized that behavioral adaptation effects in the cross-adaptation condition indicate the sensitivity of action representations to contingent actions.

In line with this hypothesis, our results show a significant cross-adaptation effect as indicated by a one-sample *t* test [$t(24) = 5.83$, Cohen's $d = 1.17$, $P < 0.001$]. In addition to sensitivity of action representations to contingent actions, action representations are more sensitive to their primary action as suggested by the significantly larger adaptation effect in the normal than in the cross-adaptation condition [paired *t* test: $t(24) = 6.26$, Cohen's $d = 1.48$, $P < 0.001$] (Fig. 2*B*). Overall, these results provide strong evidence for individual action representations being also partially activated by their contingent actions—ones that naturally occur in spatiotemporal proximity during a social interaction.

The cross-adaptation effects are particularly surprising as the adaptor and test stimuli were perceptually very dissimilar. To further validate that the cross-adaptation effect is not merely a signature of low-level adaptation (e.g., to local joint angle changes), we followed up with a control study (experiment 2). We assessed the contribution of the lower-level adaptation effects in the cross-adaptation condition. Specifically, we generated a scrambled action stimulus by remapping joint angle movements of the arms onto the legs and vice versa. This impaired the holistic percept of the action while retaining local joint angle movements and leaving the overall available movement information largely unaffected (Fig. 3*A* and Movie S3). In experiment 2, we used the four scrambled actions as an additional set of adaptors, together with the four nonscrambled adaptors and two test morphs of experiment 1. We reasoned that, if low-level visual cues were the sole contributors to the cross-adaptation effect, then adaptation to scrambled actions should lead to very similar cross-adaptation effects as adaptation to nonscrambled actions. To this end, in experiment 2 participants (*n* = 25) were adapted with non-scrambled and scrambled actions as adaptors in both the cross-adaptation and the normal condition and always probed with nonscrambled morphed test stimuli. The most important finding of this experiment was that in the cross-adaptation condition the scrambled adaptors, which contained very similar movement cues to the nonscrambled adaptors, significantly reduced the size of the adaptation effect compared with the nonscrambled adaptors [paired *t* test $t(24) = 3.56$, Cohen's $d = 1.17$, $P = 0.001$]. Specifically, the scrambled adaptors in the cross-adaptation condition did not induce significant adaptation effects [$t(24) = 1.6$, Cohen's $d = 0.32$, $P = 0.12$; the Bayes Factor (B) is $B_{10} = 0.56$ ($r = 0.707$)]. The significant reduction of the cross-adaptation effect by scrambling demonstrates that low-level visual cues alone did not drive the

A

Throwing-giving

Catching-taking

B
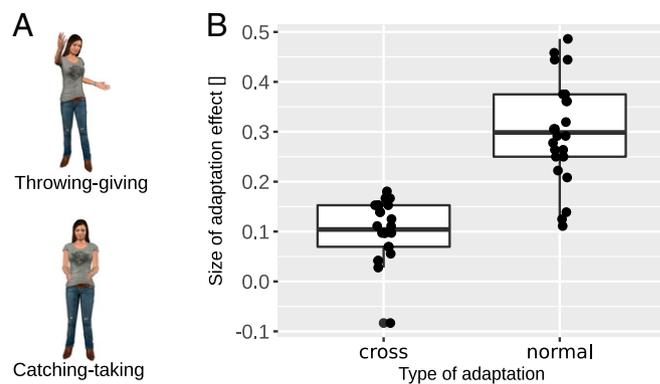
Fig. 2. (*A*) Peak frames of two exemplary action morphs used in experiment 1. (*Top*) Throwing–giving morph. (*Bottom*) Catching–taking morph. (*B*) Results of experiment 1. Boxplots of the adaptation aftereffect for cross- and normal adaptation conditions. Boxes indicate the interquartile range (IQR) with the midline in the box being the median. Whiskers mark intervals of 1.5 times the IQR ranges. Dots show individual data.
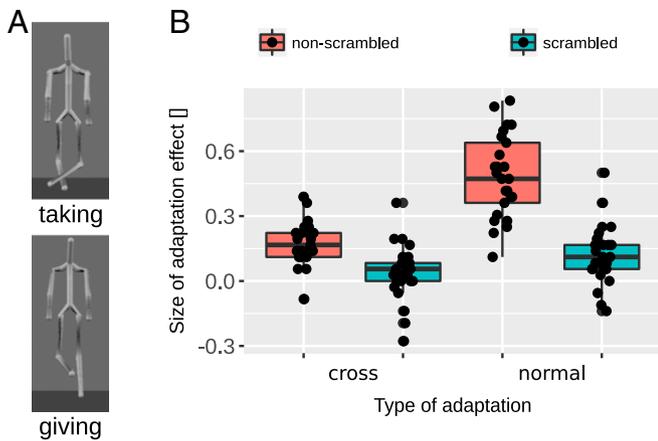
Fig. 3. (*A*) Peak frames of the two adaptors used in experiment 2. (*Top*) Scrambled taking adaptor. (*Bottom*) Scrambled giving adaptor. (*B*) Results of experiment 2. Boxplots of the adaptation aftereffect for cross- and normal adaptation conditions are shown for each scrambling condition separately (different colors). Boxes indicate the IQR with the midline in the box being the median. Whiskers mark intervals of 1.5 times the IQR ranges. Dots show individual data.

cross-adaptation effect (Fig. 3*B*). In addition, we replicated the cross-adaptation effect of the previous experiment with non-scrambled adaptors [Fig. 3*B*, $t(24) = 4.35$, Cohen's $d = 0.87$, $P < 0.001$]. Experiment 2 demonstrates that cross-adaptation effects cannot be explained by low-level visual cues alone.

Experiments 1 and 2 demonstrate the sensitivity of action representations to contingent actions. However, they do not show the selective sensitivity to contingent actions (i.e., contingency specificity). It is possible that action representations might be sensitive to any other action that is part of a social interaction. For example, the representation of a throwing action might be sensitive also to an action that is not a part of the same social interaction, e.g., dancing. To assess whether action representations exhibit some action selectivity, we examined the specificity of the cross-adaptation effect in experiment 3. We used two novel actions as adaptors that were noncontingent to the morphed actions. Namely, we used the "salsa-leader" and the "salsa-follower" actions from a salsa-dancing interaction as adaptor actions (Fig. 4*A* and Movie S4). We again probed the cross-adaptation aftereffect (in a different group of $n = 24$ participants) with the same test action morphs: throwing–giving and catching–taking (as in experiments 1 and 2). Salsa-dancing adaptor actions were noncontingent to the actions from which test stimuli were created (throwing, catching, giving, or taking actions): both salsa-leader and salsa-follower actions were unlikely to prompt a response of catching or taking, as well as these dancing actions being a response to throwing or giving actions. We predict that interaction-specific cross-adaptation effects should be modulated by how closely linked adaptation and action stimuli are. That is, salsa-dancing adaptor action should significantly reduce the cross-adaptation effect.

Because salsa dancing adaptors were noncontingent to the actions of the test stimulus, the direction of their perceptual effect on the test stimuli is not known a priori. Hence, we calculated the difference between these two noncontingent conditions post hoc in such a way that their difference resembled an action adaptation effect (i.e., their difference was positive rather than negative). By doing so, we chose the most conservative approach that is least likely to produce a significant reduction of the cross-adaptation effect in the noncontingent compared with the contingent condition. In addition, we also replicated the cross-adaptation with the contingent actions as adaptors from experiment 1. The noncontingent adaptors significantly reduced the cross-adaptation effect compared with the contingent adaptors [$t(23) = 2.41$, Cohen's $d = 0.77$, $P = 0.02$], demonstrating that action representations are sensitive to action contingency (Fig.

4*B*). The contingent condition provided evidence for a cross-adaptation effect: $t(23) = 5.26$, Cohen's $d = 1.07$, $P < 0.001$. In summary, experiment 3 demonstrates the specificity of cross-adaptation effects to the contingent action.

Does the cross-adaptation effect occur at an amodal semantic or at a perceptual level? For this, we probed the sensitivity of action representations to commonly agreed semantic labels. We chose two actions (hug and push), which we anticipated to be well known to participants. We chose another two bimanual arm actions (labeled "hand-up" and "hand-left") (Fig. 5*B* and Movies S7 and S8), which we expected to be much less well known to participants. A questionnaire, in which participants assigned these actions semantic labels, confirmed that not-well-known actions were significantly more often not assigned a semantic label compared with well-known actions: $X^2_{McNamar}(1) = 14.26$, $P > 0.001$ (*SI Appendix,* Fig. S1). Hence, if action adaptation were based on amodal semantic labels, we expect the adaptation effect to be significantly smaller for not-well-known actions compared with well-known actions. Moreover, because the motor system is often considered to be important for action recognition, we also manipulated whether actions in experiment 4 were biologically possible or impossible to execute. To obtain biologically impossible actions, we scrambled each of the four actions. We measured participants' ($n = 25$) ability to discriminate the two test morphs (well-known hug–push and the not-well-known hand-left–hand-up morphs) following scrambled and nonscrambled versions of well-known and not-well-known adaptor actions. In contrast to the previous experiments, experiment 4 probed only "normal" adaptation: we adapted only with actions that were constituent to the morph axis. Biological possibility and semantic label (well-known vs. not-well-known) were completely crossed within-subject factors. The results (Fig. 5*B*) show that action adaptation effects are modulated only by biologically possibility, $F(1, 24) = 22.6$, $P < 0.001$, not by semantic label, $F(1, 24) = 0.06$, $P = 0.8$, or by the interaction of semantic label and biological possibility, $F(1, 24) = 0.54$, $P = 0.47$. In fact, for the factor "semantic label," there is substantial evidence that the data come from the distribution assumed under the null hypothesis, $1/B_{10} = 1/0.211 = 4.74$, $r = 0.707$. The significant results demonstrate the importance of being able to motorically execute the actions for action recognition. This is in line with the prominent notion about the motor system being important for action recognition (15–17). At the same time, action representations exhibit only minor sensitivity to participants' ability to provide semantic labels for actions. This finding complements previous reports
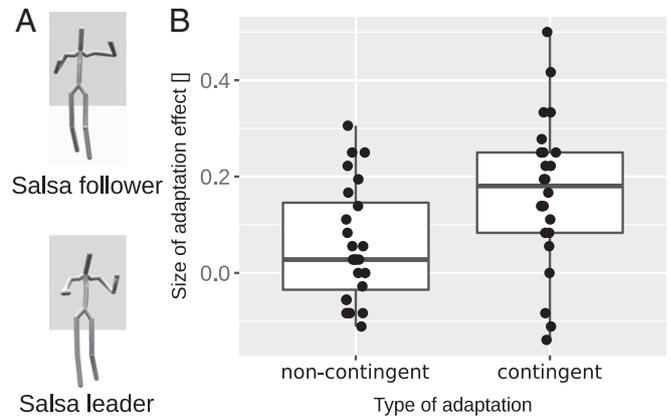
Fig. 4. (*A*) Peak frames of the two adaptors used in experiment 3. (*Top*) Salsa follower adaptor. (*Bottom*) Salsa leader adaptor. (*B*) Results of experiment 3. Boxplots of the cross-adaptation aftereffect for noncontingent (using salsa dancing as adaptors) and contingent (using throwing, giving, catching, and taking adaptors) adaptation conditions. Test stimuli were always taken from the two morph axes of Fig. 1*A*. Boxes indicate the IQR with the midline in the box being the median. Whiskers mark intervals of 1.5 times the IQR ranges. Dots show individual data.
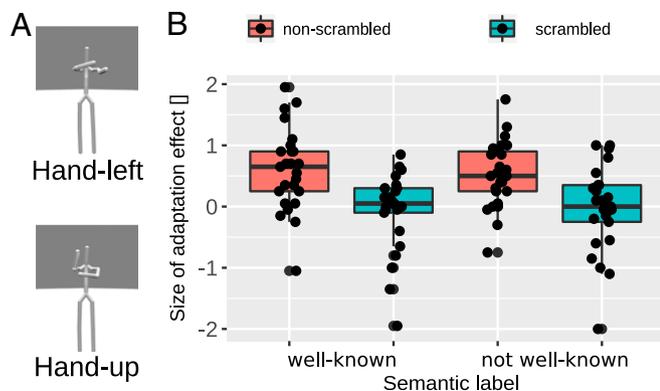
Fig. 5. (*A*) Peak frames of the two not-well-known adaptors used in experiment 4 (well-known adaptors not shown). (*Top*) Hand-left adaptor. (*Bottom*) Hand-up adaptor. (*B*) Results of experiment 4. Boxplots of the normal adaptation effect for well-known and not-well-known adaptors shown for each scrambling condition separately (different colors). Boxes indicate the IQR with the midline in the box being the median. Whiskers mark intervals of 1.5 times the IQR ranges. Dots show individual data.

(18) that demonstrate the sensitivity of action-adaptation effects to visual action information: Using action words instead of action images as adaptors abolishes the action adaptation aftereffect. Overall, the results strongly suggest that the action adaptation effect is not a purely amodal semantic adaptation effect.

Discussion

Previous studies have examined individual actions and demonstrated that the repeated presentation of an action transiently changes the percept of an action's direction (19–21) and its association with a semantic label (1, 13, 14, 18). Here we provide evidence of the representation of action contingencies between two actions in humans. In three experiments, we show a significant cross-adaptation aftereffect, demonstrating that mechanisms of action perception are sensitive to both the primary action and its contingent action. Importantly, the sensitivity to the contingent action was observed despite large visual dissimilarities between the adaptor and test stimuli. Thus, the explanation that the action representations adapt merely to the low-level visual cues (such as joint angles or low-level visual cues) in the contingent action stimuli is unlikely. A more explicit test of this explanation involved the distortion of the biological possibility of executing the action (which also leads to participants not being able to assign the meaning to these scrambled action stimuli) while retaining many of the low-level motion cues by swapping the arm and leg movements in experiment 2. In this experiment a significantly reduced cross-adaptation effect was found, although the low-level motion cues were very similar to experiment 1. Taken together, these results strongly suggest that cross-adaptation effects rely very little or not at all on overall available low-level visual information. It seems that features that are processed later in the action recognition hierarchy mediate cross-adaptation effects. Importantly, experiment 3 shows that these features are specific to the probed social interaction as the cross-adaptation effect is significantly decreased when noncontingent actions are used for adaptation. This rules out generic explanations for the cross-adaptation effect that are not social interaction-specific. In summary, we show evidence for the selective encoding of action contingencies of naturally observed interactions by action representations.

The ability to recognize action contingencies is essential for several important social cognitive functions. For example, contingent actions are better detected in noise than noncontingent actions during social interaction recognition (22–25). The ability to take advantage of action contingencies in these detection tasks seems to be related to overall social functioning. High-functioning autistic persons have been shown to take less advantage of action

contingencies to improve the detection of social interactions in noise compared with healthy controls (26). This deficit in people with autistic traits cannot be attributed to a general lack of recognizing contingencies. Children with autism focus on nonsocial contingencies when observing biological motion patterns (27), suggesting that people with autistic traits are able to detect physical contingencies. The selective impairment of using social contingencies in autistic people indicates that action contingencies are an integral part of normal social functioning.

What possible neural architecture gives rise to action recognition mechanisms being sensitive to contingent actions? The results of experiment 4 and previous results (18) show that action adaptation effects are bound to visual action information. For example, removing visual action information by using action words as adaptors, instead of action images, abolished action adaptation effects. In line with this result, experiment 4 demonstrates that an action adaptation effect appears unaltered even when participants' ability to provide semantic labels for an action deteriorates. Hence, previous and current empirical evidence suggests that adaptation is linked to visual action information and thereby to areas of visual action processing.

Visual action recognition has been associated with activation in superior temporal sulcus (STS) (28–30). Some evidence from imaging (31) and physiological (7) studies shows that visual adaptation with the action stimuli transiently changes the neural response properties in area STS. Other studies, however, show an adaptation effect in response to actions in inferior parietal lobule and inferior frontal gyrus areas (16, 32). In line with this, behavioral adaptation aftereffect and neurophysiological adaptation have been linked (33, 34). We might hypothesize that the prolonged presentation of an action transiently inhibits the neural population sensitive to the adapted action that causes the adaptation effect. When an ambiguous test action is shown that contains features of both the adapted and the nonadapted action, the nonadapted population exhibits a relatively stronger response compared with the adapted population. As a result, the ambiguous test stimulus is reported more often as the nonadapted stimulus. Within this framework the sensitivity of action recognition mechanisms to action contingencies might arise from associative learning in the following way. In social interactions, individual actions frequently co-occur in close temporal proximity, resulting in observable statistical regularities between the actions (35). These statistical regularities could result in correlated activation between the underlying neural action recognition populations. Associative learning causes this correlated activity to manifest over time by selectively strengthening the connections between these neural populations (36, 37). It could be these strengthened connections cause the activation of both the catching and the throwing action recognition processes during the observation of a catching action. Accordingly, the repeated observation of a catching action might therefore induce an adaptation effect in the throwing neural population, giving rise to the cross-adaptation effect. However, further experimentation is needed to investigate the neural architecture of the adapting contingent action representations.

We observed that cross-adaptation effects are independent of the natural temporal order in which two actions occur. Specifically, cross-adaptation effects did not statistically differ when the throwing adaptor preceded a catching–taking action morph (correct temporal order) and when the catching adaptor preceded a throwing–giving action morph (reversed temporal order). Pooling cross-adaptation effect sizes from all three experiments, we did not find significant difference between correct and reversed temporal order [$t(73) = 1.97$, Cohen's $d = 0.32$, $P = 0.05$]. The Bayes Factor for the alternative hypothesis was $B_{10} = 0.9$ ($r = 0.707$), suggesting that there is a slightly smaller, but hardly worth mentioning, likelihood in favor of the null hypothesis. Overall, our results do not allow us to draw firm conclusions about the temporal order. Thus, further studies are needed to determine whether the contingent action representations retain the temporal order of actions and, more generally, how the temporal order of social interactions is encoded in the brain.

Finally, we would like to point out the robustness of the cross-adaptation effect. We were able to replicate this effect in three independent experiments with each effect associated with a large effect size of at least Cohen's $d = 0.87$.

In conclusion, we provide evidence for the selective encoding of contingencies occurring between natural social interactions. The encoding of contingencies cannot be explained by the overall available low-level visual information. Moreover, the encoding of action contingencies is specific to the interaction. We suggest that such a mechanism is critical for humans to "see" relationships between otherwise independent social actions.

Methods

Stimuli and Apparatus. We recorded the action stimuli from real-life dyadic interactions using two motion capture suits, each equipped with 17 inertial motion trackers distributed over the whole body (MVN Motion Capture Suit from XSens). Two actors stood facing each other and carried out the actions starting from a neutral pose. We recorded three different sets of interactions between the actors: interaction 1 (giving–taking) consisted of one person giving a small bag to another person who was taking it; interaction 2 (throwing–catching) consisted of one person throwing a small bag to another person who was catching it; and interaction 3 (salsa dancing) consisted of two people salsa dancing together as partners. From these three interactions, six actions were recorded: giving, taking, throwing, catching, leading salsa, and following salsa. The recorded action stimuli contained the information about the change of 3D spatial coordinates of 22 body joints over time. All actions were processed into short movies of standardized lengths lasting 1.2 s. We also generated ambiguous action morphs between actions originating from interaction 1 or interaction 2, whereas actions were either morphed between "initiating" actions (giving–throwing morphs) or "responding" actions (catching–taking morphs). The morphs were calculated by the weighted average of local joint angles between two actions (we used the same procedure as in refs. 13 and 14). The morph levels chosen as test stimuli were determined individually in practice trials (*Procedure*). All stimuli were presented with an augmented reality setup in which participants could see the actions in 3D and were carried out by a life-size avatar (height = 1.73 m) rendered as a human female figure in experiment 1 and experiment 4 and as a stick figure for experiment 2 and experiment 3. All actions were presented with the avatar facing the viewer at a fixed distance of 2.3 m from the screen to the motion-tracked glasses. As a result, the actions were not presented within their natural social interaction context but were taken out of their spatiotemporal action context. Previous research (18) has shown that the spatiotemporal context modulates action adaptation effects in a top–down fashion. Because this top–down modulation of adaptation effect interferes with the aim of the present study to measure the sensitivity of visual action representation to contingent actions, we decided to present stimuli out of their spatiotemporal context (i.e., as single actions).

In experiment 1 with the human-like avatar, the avatar kept a neutral facial expression. The setup was programmed and controlled with the Unity game engine, and the animated avatar stimuli were acquired from Rocketbox. The stimuli were projected using a back-projection technique enabled by a Christie Mirage S+3K stereo projector with a refresh rate of 115 Hz, and all participants wore Nvidia 3D Vision Pro shutter glasses synchronized to the display to perceive the stimuli in 3D. An ART Smarttrack system was used to track the position of the head as well as the position of the hands of participants, to update the 3D visual scene in response to the viewpoint changes and enable action execution and task responses of the participants by using their hands.

Procedure.

Practice trials. At the beginning of all experiments, participants put on a motion-tracked 3D goggles and the hand-tracker. Then they were shown how to answer which actions that they perceived in the experiment by moving their tracked hands and touching one of the two virtual 3D buttons appearing midair labeled with the respective names of the actions (giving, throwing, taking, catching). Participants learned to categorize ambiguous actions to an action belonging to either interaction 1 (giving, taking) or interaction 2 (throwing, catching), depending on the morphs of ambiguous actions that were either initiating or responding actions from interaction 1 and interaction 2 (giving–throwing or catching–taking action morphs). They were able to practice this repeatedly while different morph weights between the actions were presented in ascending and descending manners to determine the point of overall ambiguous perception, which corresponds to the point of subjective equality, for each participant. All morph levels were presented twice.

Baseline condition. After the morph weight was determined in the practice trials, participants were presented with three repetitions of each test morph stimuli in the absence of adaptor stimuli (baseline). The participants who showed inconsistent (e.g., were responding with the same category to all stimuli) responses in these three iterations of the same action morphs were considered as ineligible for the experiment and excluded before participation in the main experimental phase (up to three participants were considered ineligible in experiments 1 and 2). We determined the baseline perception for each action and for each participant. Once determined, the morph weights in the test stimuli were kept identical for each participant across all experimental blocks during the main experimental phase. The total number of trials in the baseline condition consisted of 42 trials (two action morphs × seven morph levels × three repetitions) for each participant.

The same baseline and practice phases were used in all three experiments as a preliminary step to familiarize participants with the setup.

(Main) adaptation experiment 1. The main adaptation experiment consisted of eight experimental blocks, a fully crossed design with the factors adaptor stimuli (4×) and test morphs (2×). The order of experimental blocks was completely balanced across all participants. The adaptor stimuli consisted of the four recorded actions (giving, throwing, taking, and catching), each presented in separate experimental blocks. The test stimuli were morphed actions either between initiating actions or between responding actions from interaction 1 and interaction 2 (giving–throwing or catching–taking action morphs). The test stimuli consisted of a set of seven different morph weights with equal morph distances between each other, and each morphed action was shown three times in each adaptor condition, resulting in 21 test stimuli presentations in each experimental block. The order of test stimuli presentation within each experimental block was completely randomized. Each experimental block of trials consisted of an initial adaptation phase where the same adaptor stimulus was repeatedly presented 30 times with an interstimulus interval (ISI) of 250 ms. The main action categorization phase directly followed, where the adaptor stimulus was repeatedly presented four times (same ISI of 250 ms) before each trial, and then the test stimuli appeared for a two-alternative forced choice task where participants had to judge which action they perceived (e.g., "Did you see giving or throwing?"). The ISI between the adaptor stimuli and test stimuli was 400 ms. The next trial started as soon as the participants recorded their answer by moving their tracked hands toward the virtual button. The total number of trials for the main adaptation experiment 1 was 168 trials (eight experimental blocks × seven morph levels × three repetitions of test stimuli) for each participant. Participants took about 60 min to finish the whole experiment, taking ~5 min per experimental block.

Adaptation experiment 2. The second adaptation experiment consisted of 16 experimental blocks, a fully crossed design with the factors adaptor action (4×), test morphs (2×), and whether the adaptor action was scrambled or not (2×). The order of experimental blocks was completely balanced across all participants. The adaptor stimuli consisted of the four recorded actions (giving, throwing, taking, and catching), each presented in a separate experimental block, and four of the same actions in which the joint angles were permuted between the arms and the legs (scrambled). The test stimuli were the same as in experiment 1: the morphed actions either between initiating actions or between responding actions from interaction 1 and interaction 2 (giving–throwing or catching–taking action morphs). The total number of trials for the main adaptation experiment 2 was 288 trials (16 experimental blocks × 6 morph levels × 3 repetitions of test stimuli) for each participant. Participants took about 115 min to finish the whole experiment, taking ~5 min per experimental block.

Adaptation experiment 3. The third adaptation experiment consisted of 12 experimental blocks, a fully crossed design with the factors adaptor action (6×) and test morphs (2×). The order of experimental blocks was counterbalanced across all participants. The adaptor stimuli consisted of the six recorded actions (giving, throwing, taking, catching, leading salsa, following salsa), each presented in a separate experimental block. The test stimuli were the same as in experiments 1 and 2: the morphed actions either between initiating or between responding actions from interaction 1 and interaction 2 (giving–throwing or catching–taking action morphs). The total number of trials for the main adaptation experiment 3 was 216 trials (12 experimental blocks × 6 morph levels × 3 repetitions of test stimuli) for each participant. Participants took about 80 min to finish the whole experiment, taking ~5 min per experimental block.

Adaptation experiment 4. This experiment was a preregistered study within the Open Science Framework (https://osf.io/hnufe). The experiment consisted of 12 experimental blocks, which resulted from completely crossing the following three within-subject factors: three adaptor conditions (baseline, adaptor action 1, adaptor action 2), two semantic label (well-known and not-well-known actions) conditions, and two biological possibility conditions (scrambled vs.

nonscrambling the actions). The testing order of experimental blocks was randomized. The adaptor stimuli consisted of two recorded well-known actions (hug and push) and two physically possible but much less known actions (hand-left and hand-up). For the biologically impossible actions, we scrambled these four actions by swapping joint angle movements between arms and legs. There were four test stimuli generated by the joint angle morphing, giving four morph axes: "hug–push scrambled," "hug–push nonscrambled," "hand-left–hand-up scrambled," and "hand-left–hand-up nonscrambled." Each of the four morph levels was probed five times. Hence, the total number of trials in experiment 4 was 240 trials (12 experimental blocks × 5 morph levels × 4 repetitions of test stimuli) for each participant. Participants took about 60 min to finish the whole experiment, taking ∼5 min per experimental block.

Assessment of semantic action labels. After finishing all blocks in experiment 4, participants were given a questionnaire, which assessed whether each of the eight adaptor action stimuli could be labeled with a semantic meaning. Next to an image of the last frame of the action (showing the action at its peak), participants answered the following four questions: "Do you know the meaning of the action (the interacting object might be missing)? If you know it, please name the action with one word (do not describe the action); otherwise, write "no" and "How confident are you on a scale from 1 (not at all) to 10 (completely)?". We coded the responses to the first question to represent the verb in the infinitive form and to remove the noun from the verb phrase, which was otherwise left unchanged. *SI Appendix*, Fig. S1, shows distributions of the coded responses of each of the action stimuli of experiment 4. For hug or push the majority of participants (81 and 85.7%, respectively) identified the action correctly. For every one of the four scrambled action stimuli, the majority of participants responded "no" ("60% for the scrambled push, 63.2% for the scrambled hug, 90% for the scrambled hand-up, 61.9% for the scrambled hand-left), according to the questionnaire guideline to respond "no" in case they did not know the meaning of the action. For (nonscrambled) hand-left and hand-up stimuli, the most prominent answer was "no" (23.8% for the hand-up and 38.1% for hand-left), and none of the other response categories gathered more than 14.3% for hand-up and 4.76% for hand-left. This demonstrates that only hug and push actions were associated with a commonly accepted meaning.

Statistical Analysis. Statistical analysis was performed using R version 3.4 (38), RStudio version 1.1.4 (39). We used ggplot2 (40) version 2.2 and plotly (41) to generate the statistical plots; Bayes factors were computed using BayesFactor R package version 0.9.2, and calculations in analysis of variance were done using ez R package version 4.4.

Participants. Twenty-five (*n* = 25) volunteers participated in experiment 1, a distinct group of 25 (*n* = 25) volunteers participated in experiment 2, a distinct group of 24 (*n* = 24) volunteers participated in experiment 3, and a distinct group of 25 (*n* = 25) volunteers participated in experiment 4. All participants were compensated with 8 € per hour. After the experiments, they were debriefed and informed about the study. Exclusion criteria applied to participants who could not perceive the action stimuli with the 3D goggles, who showed inconsistent responses after three iterations in the initial process determining the point of overall ambiguous perception for the two morphed actions, or who reported to have had extensive training in sports involving throwing or catching actions (e.g., experts in baseball, basketball, juggling) or in salsa dancing (in the case of experiment 3). In Experiment 1, three participants met the exclusion criteria; in experiment 2, two participants met the exclusion criteria; and in experiments 3 and 4, no participants met the exclusion criteria.

Consent. Psychophysical experiments were performed with informed consent of participants. All participants were informed about the purpose of the experiment before signing an informed consent. All participants were naive concerning the hypotheses of the experiment. The experiment was conducted in line with the Declaration of Helsinki and was approved by the ethics board of the University of Tübingen (Germany).

1. de la Rosa S, Ekramnia M, Bülthoff HH (2016) Action recognition and movement direction discrimination tasks are associated with different adaptation patterns. *Front Hum Neurosci* 10:56.
2. Clifford CW, et al. (2007) Visual adaptation: Neural, psychological and computational aspects. *Vision Res* 47:3125–3131.
3. Lawson RP, Clifford CW, Calder AJ (2009) About turn: The visual representation of human body orientation revealed by adaptation. *Psychol Sci* 20:363–371.
4. Bekesy Gv (1929) Zur Theorie des Hörens: Über die Bestimmung des einem reinen Tonempfinden entsprechenden Erregungsgebietes der Basilarmembran vermittelst Ermüdungserscheinungen [Regarding a theory of hearing: About the determination of excitation area of the basilar membrane regarding pure tone listening]. *Phys Z* 30: 115–125. German.
5. Grill-Spector K, Malach R (2001) fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107:293–321.
6. Webster MA (2011) Adaptation and visual coding. *J Vis* 11:3.
7. Kuravi P, Caggiano V, Giese M, Vogels R (2016) Repetition suppression for visual actions in the macaque superior temporal sulcus. *J Neurophysiol* 115:1324–1337.
8. Kuravi P, Vogels R (2017) Effect of adapter duration on repetition suppression in inferior temporal cortex. *Sci Rep* 7:3162.
9. Vogels R (2016) Sources of adaptation of inferior temporal cortical responses. *Cortex* 80:185–195.
10. Kaliukhovich DA, De Baene W, Vogels R (2013) Effect of adaptation on object representation accuracy in macaque inferior temporal cortex. *J Cogn Neurosci* 25:777–789.
11. Kaliukhovich DA, Vogels R (2012) Stimulus repetition affects both strength and synchrony of macaque inferior temporal cortical activity. *J Neurophysiol* 107:3509–3527.
12. Kaliukhovich DA, Vogels R (2011) Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cereb Cortex* 21:1547–1558.
13. de la Rosa S, Ferstl Y, Bülthoff HH (2016) Visual adaptation dominates bimodal visual-motor action adaptation. *Sci Rep* 6:23829.
14. Ferstl Y, Bülthoff H, de la Rosa S (2017) Action recognition is sensitive to the identity of the actor. *Cognition* 166:201–206.
15. Rizzolatti G, Craighero L (2004) The mirror-neuron system. *Annu Rev Neurosci* 27:169–192.
16. Kilner JM, Neal A, Weiskopf N, Friston KJ, Frith CD (2009) Evidence of mirror neurons in human inferior frontal gyrus. *J Neurosci* 29:10153–10159.
17. Becchio C, et al. (2012) Social grasping: From mirroring to mentalizing. *Neuroimage* 61:240–248.
18. de la Rosa S, Streuber S, Giese M, Bülthoff HH, Curio C (2014) Putting actions in context: Visual action adaptation aftereffects are modulated by social contexts. *PLoS One* 9:e86502.
19. Liepelt R, Prinz W, Brass M (2010) When do we simulate non-human agents? Dissociating communicative and non-communicative actions. *Cognition* 115:426–434.
20. Theusner S, de Lussanet MH, Lappe M (2011) Adaptation to biological motion leads to a motion and a form aftereffect. *Atten Percept Psychophys* 73:1843–1855.
21. Sartori L, Straulino E, Castiello U (2011) How objects are grasped: The interplay between affordances and end-goals. *PLoS One* 6:e25203.
22. Manera V, et al. (2012) Are you approaching me? Motor execution influences perceived action orientation. *PLoS One* 7:e37514.
23. Manera V, Schouten B, Verfaillie K, Becchio C (2013) Time will show: Real time predictions during interpersonal action perception. *PLoS One* 8:e54949.
24. Manera V, Becchio C, Schouten B, Bara BG, Verfaillie K (2011) Communicative interactions improve visual detection of biological motion. *PLoS One* 6:e14594.
25. Neri P, Luu JY, Levi DM (2006) Meaningful interactions can enhance visual discrimination of human agents. *Nat Neurosci* 9:1186–1192.
26. von der Luhe T, et al. (2016) Interpersonal predictive coding, not action perception, is impaired in autism. *Philos Trans R Soc Lond B Biol Sci* 371:20150373.
27. Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W (2009) Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* 459:257–261.
28. Vaina LM, Solomon J, Chowdhury S, Sinha P, Belliveau JW (2001) Functional neuroanatomy of biological motion perception in humans. *Proc Natl Acad Sci USA* 98:11656–11661.
29. Grossman ED, Jardine NL, Pyles JA (2010) fMR-adaptation reveals invariant coding of biological motion on the human STS. *Front Hum Neurosci* 4:15.
30. Pelphrey KA, Morris JP, Michelich CR, Allison T, McCarthy G (2005) Functional anatomy of biological motion perception in posterior temporal cortex: An FMRI study of eye, mouth and hand movements. *Cereb Cortex* 15:1866–1876.
31. Jastorff J, Kourtzi Z, Giese MA (2009) Visual learning shapes the processing of complex movement stimuli in the human brain. *J Neurosci* 29:14026–14038.
32. Hamilton AF, Grafton ST (2008) Action outcomes are represented in human inferior frontoparietal cortex. *Cereb Cortex* 18:1160–1168.
33. Thurman SM, van Boxtel JJ, Monti MM, Chiang JN, Lu H (2016) Neural adaptation in pSTS correlates with perceptual aftereffects to biological motion and with autistic traits. *Neuroimage* 136:149–161.
34. Barraclough NE, Keith RH, Xiao D, Oram MW, Perrett DI (2009) Visual adaptation to goal-directed hand actions. *J Cogn Neurosci* 21:1806–1820.
35. Heyes C (2010) Mesmerising mirror neurons. *Neuroimage* 51:789–791.
36. Cooper RP, Cook R, Dickinson A, Heyes CM (2013) Associative (not Hebbian) learning and the mirror neuron system. *Neurosci Lett* 540:28–36.
37. Keysers C, Perrett DI (2004) Demystifying social cognition: A Hebbian perspective. *Trends Cogn Sci* 8:501–507.
38. R Core Team (2014) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna), Version 3.4.0. Available at https://www.R-project.org. Accessed June 18, 2018.
39. RStudio Team (2016) RStudio: Integrated Development for R (RStudio, Inc., Boston), Version 1.1.447. Available at www.rstudio.org. Accessed June 18, 2018.
40. Wickham H (2009) ggplot2: *Elegant Graphics for Data Analysis* (Springer, New York).
41. Plotly Technologies Inc. (2015) *Collaborative Data Science* (Plotly Technologies, Montreal).