Report

Current Biology

Mirror Neurons in Monkey Premotor Area F5 Show Tuning for Critical Features of Visual Causality Perception

Highlights

- Mirror neurons respond to abstract visual stimuli, suggesting causation
- Responses to natural hand actions and to abstract stimuli are highly similar
- Manipulations destroying causality perception in humans modulate visual responses

Authors

Vittorio Caggiano, Falk Fleischer, Joern K. Pomper, Martin A. Giese, Peter Thier

Correspondence

caggiano@gmail.com (V.C.), martin.giese@uni-tuebingen.de (M.A.G.)

In Brief

Humans perceive "causality" from abstract motion stimuli, e.g., when one disc bumps into another one. Caggiano et al. show that response patterns of mirror neurons during the perception of abstract causality stimuli resemble the ones to natural hand actions. Representations for action and causality perception thus overlap at the single-cell level.

Caggiano et al., 2016, Current Biology *26*, 3077–3082
November 21, 2016
http://dx.doi.org/10.1016/j.cub.2016.10.007

CellPress

Current Biology
Report

Mirror Neurons in Monkey Premotor Area F5 Show Tuning for Critical Features of Visual Causality Perception

Vittorio Caggiano,[1,3,5,6,*] Falk Fleischer,[2,3] Joern K. Pomper,[1,3] Martin A. Giese,[2,4,*] and Peter Thier[1,4]
[1]Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany
[2]Section for Computational Sensomotorics, Department of Cognitive Neurology, Centre for Integrative Neuroscience and Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany
[3]Co-first author
[4]Co-senior author
[5]Present address: Computational Biology Center, IBM T.J. Watson Research Center, 1101 Kitchawan Road, Route 134, Yorktown Heights, NY 10598, USA
[6]Lead Contact
*Correspondence: caggiano@gmail.com (V.C.), martin.giese@uni-tuebingen.de (M.A.G.)
http://dx.doi.org/10.1016/j.cub.2016.10.007

SUMMARY

Humans derive causality judgments reliably from highly abstract stimuli, such as moving discs that bump into each other [1]. This fascinating visual capability emerges gradually during human development [2], perhaps as consequence of sensorimotor experience [3]. Human functional imaging studies suggest an involvement of the "action observation network" in the processing of such stimuli [4, 5]. In addition, theoretical studies suggest a link between the computational mechanisms of action and causality perception [6, 7], consistent with the fact that both functions require an analysis of sequences of spatiotemporal relationships between interacting stimulus elements. Single-cell correlates of the perception of causality are completely unknown. In order to find such neural correlates, we investigated the responses of "mirror neurons" in macaque premotor area F5 [8, 9]. These neurons respond during the observation as well as during the execution of actions and show interesting invariances, e.g., with respect to the stimulus view [10], occlusions [11], or whether an action is really executed or suppressed [12]. We investigated the spatiotemporal properties of the visual responses of mirror neurons to naturalistic hand action stimuli and to abstract stimuli, which specified the same causal relationships. We found a high degree of generalization between these two stimulus classes. In addition, many features that strongly reduced the similarity of the response patterns coincided with the ones that also destroy the perception of causality in humans. This implies an overlap of neural structures involved in the processing of actions and the visual perception of causality at the single-cell level.

RESULTS

In order to study the responses of mirror neurons to abstract stimuli that suggest causal relationships, we proceeded in two steps. First (experiment 1), we investigated the spatiotemporal structure of the responses of mirror neurons to filmed naturalistic hand-object interactions (called "naturalistic stimuli" in the following). In a second step (experiment 2), we investigated how these responses changed if the stimuli were replaced by abstract patterns (interacting discs) that specified the same causal relationships. In addition, we tested various control stimuli that lacked different features of the original abstract stimuli in order to determine the stimulus properties that are critical for inducing similar response patterns in mirror neurons. These stimulus classes were presented in a randomly interleaved fashion, where neurons tested in experiment 2 were a subset of the ones in experiment 1.

Following our previous work [10, 13], neurons were characterized as "mirror neurons" when they discharged during both the execution of goal-directed motor acts and during the observation of action movies (see the Supplemental Information for a detailed description). Such movie stimuli (unlike acts executed by the experimenter in front of the monkey) permit an exact control of the timing and spatiotemporal structure. Monkeys were rewarded when they kept their eyes within the stimulus window, independent of the presented stimulus. This experimental design prevented stimulus-specific learning effects. In total, we recorded from 1,126 neurons in area F5 with motor act-related responses (619 for monkey E and 507 for monkey P). Out of these 489 neurons, 43% were typical mirror neurons (120 neurons from monkey E and 88 from monkey P). (See the Supplemental Information for further details.)

Neural Responses to Naturalistic Hand Actions

In the first experiment, we tested each neuron with movies showing a naturalistic hand action ("grasping" of a piece of food), either with normal or reversed temporal order of the frames (Figure 1A and Movie S1). Temporally reversed grasping looks like the placing of an object with subsequent removal of the hand, eliciting virtually identical neural responses as real placing

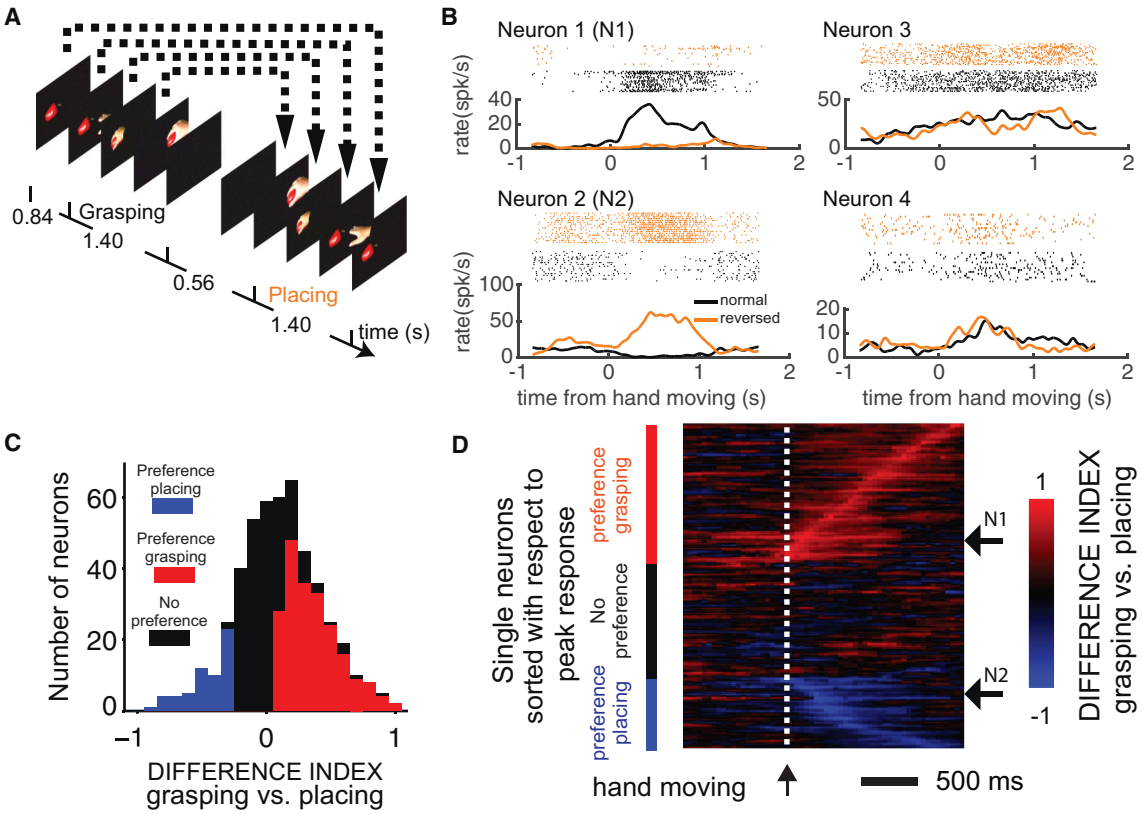

Figure 1. Neural Encoding of Observed Naturalistic Actions in F5 Mirror Neurons

(A) Test movie (Movie S1) presenting a hand grasping an object (grasping) followed by same movie with reversed temporal order of frames (placing).

(B) Firing rates of four typical example units after temporal realignment (reversing and aligning temporal axis for placing stimulus). Neurons 1 and 2 show preferences for grasping versus placing, whereas neurons 3 and 4 do not show such a preference.

(C) Histogram of the temporal averages of the difference index (DI) for neurons with selectivity for grasping, placing, and neurons without significant preference.

(D) Time courses of the DIs for the individual neurons (with preferences for grasping, placing, and non-selective ones). Neurons are ordered along vertical axis with respect to the time points of the extrema of the DI over time (N1 and N2 indicate neurons shown in B).

See also Figures S1 and S2.

stimuli (cf. [14]). Out of the tested mirror neurons, 315 (64%) showed a significant preference for one of the two sequential frame orders (sign-rank test for average firing rate; $p < 0.05$), where 208 (42%) preferred grasping and 107 (22%) "placing" actions. Figure 1B shows example units responding preferentially to one temporal order (neurons 1 and 2) and neurons without order preference (neurons 3 and 4). Normalized population responses to these stimuli are shown in Figure S1.

In order to characterize the temporal sequence selectivity of the individual neurons more accurately, we compared the responses to the same movie frames in the forward sequence (grasping) and the time-reversed sequence (placing). If responses were independent of the temporal order, the response traces to grasping and the time-reversed response traces to placing would be highly similar, because both stimuli present the same movie frames just in opposite order. The similarity of these response traces was quantified by defining a difference index (DI), which is given by the difference of the responses in corresponding time windows, normalized by their sum (a correlation analysis in the Supplemental Information supports the efficiency of this analysis). The DI was computed for overlapping time windows with a duration of 200 ms and sampling steps of 20 ms. For neurons that respond exclusively to grasping, but not to placing (within a particular time interval), the DI would be 1. Likewise, if a neuron responds exclusively to placing in a particular time interval, but not to grasping, the DI is −1. For neurons that are not sequence selective, the DI is 0.

Figure 1C shows a histogram of the time averages of the DIs over all tested mirror neurons. The distribution spans over the whole range of DIs, from neurons with a preference for grasping characterized by positive DIs significantly deviating from zero (median 0.33; sign-rank test; $p < 0.001$) to neurons with preference for placing identified by significant negative average DIs (median −0.26; $p < 0.001$). For neurons that did not exhibit a significant preference for either of the two stimuli, the population DI remained very close to 0 (median 0.02; larger than 0 with $p = 0.009$).

The time courses of the DIs for the individual neurons are shown in Figure 1D (neurons with preference for grasping and placing in the upper and lower parts, respectively, and neurons without such preference in the middle part of the panel). Most neurons with preference for grasping or placing showed only

one dominant extremum of the DI over time, often close to 1 or −1. This means that many of these neurons were maximally active only for a limited time interval, responding more weakly when the same frame was embedded in a stimulus sequence with opposite temporal order. Neurons are ordered along the vertical axis according to the time points of the extrema of the DI. These time points span over the whole time interval of the action from the start of the hand movement with a relatively uniform distribution, especially for the grasping-selective neurons. The observed spatiotemporal activation pattern resembles an activation maximum that travels over the population. (See the Supplemental Information for further discussion.)

It is important to verify that the observed spatiotemporal responses do not just reflect simple low-level stimulus features (e.g., local motion or shape features) that covary with manipulations affecting causality. This possibility is ruled out by two additional control experiments that are described in detail in the Supplemental Information. During the presentation of sequences with multiple actions (e.g., grasping-placing versus placing-grasping or three-action sequences), the responses of many neurons to the individual action changed dramatically when the sequence order of the actions was altered (Figures S2A–S2C and S2D–S2G). This behavior is definitely incompatible with trivial explanations in terms of a tuning to instantaneous low-level features and indicates an involvement of higher representations that integrate information over multiple actions [9, 15] (cf. Supplemental Information).

Responses to Causal and Non-causal Abstract Stimuli

After establishing the spatiotemporal response patterns to naturalistic action stimuli in experiment 1, we compared in experiment 2 the responses to movies of naturalistic hand actions and to abstract stimuli, which specified the same causal relationships. The abstract stimuli consisted of two moving discs whose positions matched the centers of the hand and the goal object in the naturalistic action movies (see Movie S1 and the Supplemental Information for details). These stimuli lack important action-specific features, such as the shape and texture of the hand and the grasped object, and the grip affordances of the latter. If mirror neurons were mainly tuned to the causal structure of the interaction between effector and object rather than detailed action-specific features, one would expect them to respond in a very similar manner to naturalistic and such abstract stimuli.

We compared the responses of mirror neurons to three classes of stimuli (cf. Figure 2A and Movie S1): naturalistic hand actions (n), movies in which the hand was replaced by a disc (h), and abstract stimuli in which both hand and grasped object were replaced by discs (c). Figure 2B shows the activities of two typical neurons that exhibit remarkably similar responses to all three types of stimuli. For more than 73% of the recorded mirror neurons (total tested 232), the responses (average discharge rates during grasping and placing) in the three conditions were not significantly different. Comparing the responses of individual cells, for the stimulus types (n) and (h), only 40 (17%) of the mirror neurons showed significantly different responses to grasping and 41 (18%) to placing. Comparing the responses for stimulus types (n) and (c), we found that 62 (27%) of the neurons showed significantly different responses to grasping and 41 (18%) to placing ($p < 0.05$; U test). Finally, comparing the responses between types (h) and (c), only nine (4%) neurons showed significant differences for grasping and six (3%) for placing. The response profiles are highly similar, as indicated by their highly significant cross-correlation coefficients (for time lag zero; $r^2 = 0.67$ (n) versus (h); $r^2 = 0.63$ (n) versus (c); $r^2 = 0.73$ (h) versus (c); $p < 0.001$). Also, the average activities for grasping and placing stimuli are highly similar across stimulus types (pairwise Spearman correlations $r^2 \geq 0.88$ for grasping and $r^2 \geq 0.76$ for placing; $p < 0.001$). Summarizing, these results show that mirror neurons indeed show a high degree of generalization between naturalistic actions and abstract stimuli that specify the same causal relationships.

This striking similarity between the responses to the two very different stimulus classes was also found when comparing the time courses of the DI, computed in the same manner as for Figure 1D. The time courses for the DIs of all tested neurons (Figure 2C) are extremely similar for the stimulus types (n) and (c), where the ordering of the neurons along the vertical axis is the same for both panels (derived from the timing of the extrema for the naturalistic stimulus), in spite of the strong differences in terms of the underlying visual features. This similarity is confirmed by highly significant cross-correlations (time lag zero; $p < 0.001$) between the temporal profiles of the DI for the three stimulus classes (medians of the cross-correlation coefficients: $r^2 = 0.48$ comparing stimulus types (n) and (h); $r^2 = 0.41$ comparing types (h) and (c); $r^2 = 0.49$ comparing types (n) and (c)). The high similarity is also supported by the high correlation of the time averages of the DI of the individual neurons for the different stimulus types ($r^2 = 0.63$, $p < 0.05$ for types (n) and (h); $r^2 = 0.77$, $p < 0.001$ for types (h) and (c); $r^2 = 0.68$, $p < 0.001$ for types (n) and (c); see also Figure 2D). Summarizing, these results demonstrate that F5 mirror neurons generalize from naturalistic hand actions to abstract stimuli that specify similar causal relationships between the stimulus elements, at the level of individual neurons as well as with respect to the spatiotemporal activation patterns of the whole population. At the same time, the observed highly structured spatiotemporal activation patterns cannot be explained by unspecific tuning to low-level visual features.

The remarkable level of generalization to highly abstract stimuli raises the question which features are critical for the activation of mirror neurons, and more specifically, whether the relevant features coincide with the ones that are known to be critical for the perception of causality in humans [1]. To answer this question, we tested a set of control stimuli that lacked various features of the naturalistic stimuli that were either important or unimportant for the causal interactions of the stimulus elements (see Movie S1). Manipulations that did not affect causality included the replacement of hand and object by discs and the exchange of the colors of these discs. A larger number of manipulations affected the causal relationship between the stimulus elements (modification of the trajectories, introduction of a spatial gap between the discs, removing one element, and repetition of reaching without the subsequent effect of the hand on the object). Many of these manipulations coincide with the ones that Michotte demonstrated to impair the perception of causality in humans. An overview of these stimuli is given in Figure 3. (See the

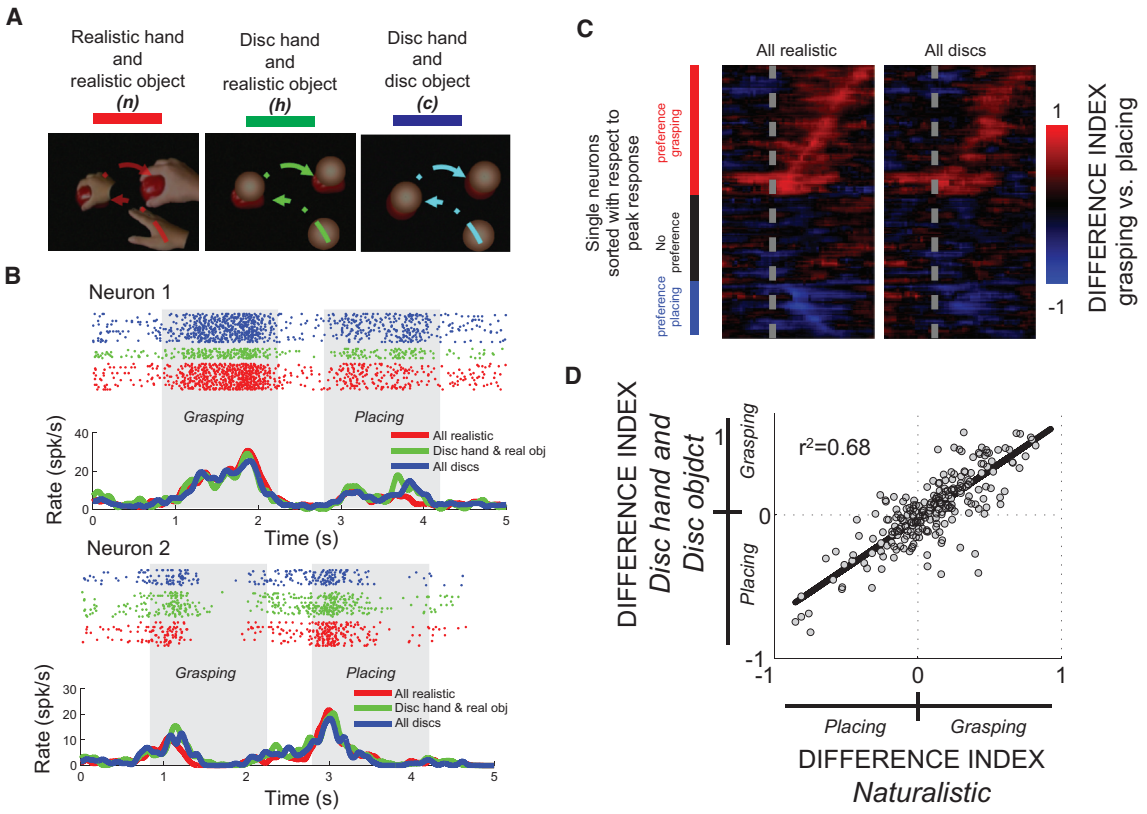

Figure 2. Responses to Abstract Causality Stimuli versus Naturalistic Actions

(A) Naturalistic stimuli (n) and abstract stimuli, generated by replacing hand (h) or hand and object by discs (c) (see the Supplemental Information and Movie S1).

(B) Spike train of two example neurons for the three stimulus types.

(C) Time courses of the DIs for the individual neurons for natural and abstract stimuli. Ordering of the neurons along the vertical axis is the same in both panels (cf. Figure 1D).

(D) Difference indices for stimulus classes (n) and (c) (correlation $r^2 = 0.68$; $p < 0.001$).

See also Movies S2 and S3.

Supplemental Information for a detailed explanation and Movies S1 and S2 for illustrations.)

The similarities between the response patterns evoked by the control stimuli and the naturalistic stimuli were quantified in two different ways. First, we computed the correlations between time-averaged DIs over all neurons between the responses. These correlations are shown in Figure 3 (middle panel). The first three control conditions that specified stimuli with similar causal relationships between stimulus elements but varying shape and color resulted in high correlations of the neural responses with the ones for the naturalistic stimulus (correlations $r^2 > 0.64$; $p < 0.001$). In contrast, for the seven control stimuli with changes that affected the causal structure, the correlations were significantly lower ($r^2 < 0.4$; $p < 0.05$; correlation differences significant with $p < 0.01$; Fisher z transform).

A completely consistent picture arose based on a second, quite different approach for data analysis. The instantaneous firing rates of all tested mirror neurons were subjected to a principal-component analysis (PCA) to construct a neural state space with reduced dimensionality (M = 20). The response traces of the population correspond to "neural trajectories" in this reduced-state space. We computed normalized distances between the neural trajectories for the naturalistic stimulus and the control conditions (see Figure S3 and Supplemental Information for further details). The computed distances d_N between the neural trajectories were completely consistent with the previous analysis (Figure 3, right panel): for the three control conditions with similar causality structure, the distances between the neural trajectories for the control and the naturalistic stimuli were significantly lower ($d_N < 0.24$) than the distances for the seven control conditions, which changed the causal structure of the stimulus ($d_N > 0.35$; $p < 0.05$; see bootstrap analysis in the Supplemental Information). The population responses for stimuli with changes in the causal structure are thus much more dissimilar to the one for the naturalistic stimuli than the ones for controls with changes that leave the causal structure intact. In other words, mirror neurons do not generalize with respect to all properties of the artificial stimuli and show a high selectivity for features that change causal relationships.

DISCUSSION

We reported experiments that provide the first evidence for individual neurons whose responses reflect properties of the "visual

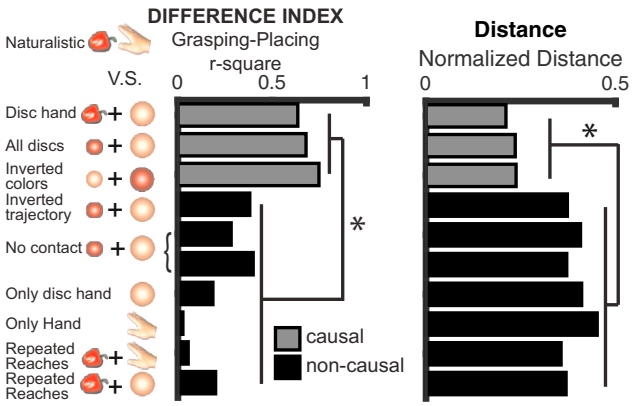

Figure 3. Responses to Causal and Non-causal Stimuli versus Naturalistic Actions

Correlations between DIs for naturalistic and control stimuli (left) and distances between the neural trajectories for the naturalistic stimulus and different control stimuli (right). The first three control stimuli (gray bars) leave the causal relationships between the stimulus elements intact. The last seven control conditions (black bars) reduce the impression of causality in humans. Asterisks indicate significant differences between stimulus groups; $p < 0.05$ (see the main text, Figure S3, and the Supplemental Information for further details). See also Movies S1, S2, and S3.

perception of causality," as defined in [1] for humans. We found that mirror neurons in macaque premotor area F5 responded to abstract stimuli, consisting of moving discs that specified similar causal relationships as naturalistic hand actions, and showed a remarkable degree of generalization between naturalistic and such abstract stimuli. The observed spatiotemporal population responses can be characterized by an ordered sequential activation of sequence-selective neurons. Control experiments showed that this generalization is highly specific and breaks down for stimulus manipulations that impair the perception of causal relationships between the interacting stimulus elements in humans. Importantly, the observed responses cannot be explained by simple low-level stimulus properties, such as local motion or form features. The causal relationships between the interacting stimulus elements are defined by specific visual features that were narrowed down by the control experiment presented in Figure 3.

Our study supports a direct relationship between the neural encoding of actions and the visual perception of causality, consistent with previous imaging results in humans (e.g., [4, 5]) and recent theoretical work [6, 7], and it establishes the existence of neurons that respond selectively to both stimulus classes. However, it leaves several fundamental questions unresolved.

First, we tested only neurons in area F5, and it seems likely that neurons in other high-level areas, such as the prefrontal cortex, might also exhibit similar invariance properties. A complete characterization of all cortical structures of the monkey's brain that show selectivity for the perception of causality would require an exhaustive screening, which exceeds the scope of this study.

Second, our study is not suitable to decide whether the investigated representations in area F5 form a crucial step in the detection of perceptual causality. Future experiments, e.g., training the animals to categorize causal versus non-causal stimuli combined with an inactivation of area F5, might help to provide answers to this question.

Third, the investigated form of causal relationships is an extremely simple one, and it remains to be studied whether, from the underlying neural mechanisms, also something might be learned about other, more advanced forms of causal reasoning in humans, as studied in cognitive science (e.g., [16]). (A more extended discussion about definition and psychological theories about causality is given in the Supplemental Information).

Fourth, addressing the question how causality perception in humans and monkeys are related, the Supplemental Information discusses evidence supporting that rhesus monkeys most likely show very simple forms of causality detection, whereas higher forms of causal reasoning quite certainly are not present. The question of whether monkeys' visual perception really matches the one of humans is impossible to test because this would require the assessment of subjective causality impressions of the animals, which would require verbal reporting. This shortcoming is shared between our study of monkeys and studies in developmental psychology addressing preverbal children.

In spite of these limitations, which motivate future studies, the hypothesis that some higher human cognitive capabilities might have evolved by generalization from very elementary neural processes in action recognition seems intriguing.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, one table, and two movies and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.10.007.

AUTHOR CONTRIBUTIONS

V.C., F.F., M.G., and P.T. conceived and designed the experiments. V.C. performed the experiments with contributions from F.F. and J.K.P. V.C. and M.G. analyzed data. All authors discussed results. V.C., M.G., and P.T. wrote the paper. P.T. supervised and provided continuous practical support.

ACKNOWLEDGMENTS

We thank four anonymous reviewers for excellent comments. V.C., J.K.P., and P.T. were supported by DFG GI 305/4-1. M.G. and F.F. were supported by DFG GI 305/4-1, KA 1258/15-1, EC FP7 PEOPLE-2011-ITN (ABC), ICT-2013-10/ 611909 (Koroibot), and H2020 ICT-23-2014/644727 (Cogimon).

Received: July 6, 2016
Revised: September 14, 2016
Accepted: October 6, 2016
Published: November 3, 2016

REFERENCES

1. Michotte, A.E.d. (1946). La Perception de la Causalité (Institut Supérieur de Philosophie).
2. Saxe, R., and Carey, S. (2006). The perception of causality in infancy. Acta Psychol. (Amst.) *123*, 144–165.
3. Sommerville, J.A., Hildebrand, E.A., and Crane, C.C. (2008). Experience matters: the impact of doing versus watching on infants' subsequent perception of tool-use events. Dev. Psychol. *44*, 1249–1256.
4. Blakemore, S.J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A.N., Segebarth, C., and Decety, J. (2001). How the brain perceives causality: an event-related fMRI study. Neuroreport *12*, 3741–3746.

5. Fugelsang, J.A., and Dunbar, K.N. (2005). Brain-based mechanisms underlying complex causal thinking. Neuropsychologia *43*, 1204–1213.
6. Fleischer, F., Christensen, A., Caggiano, V., Thier, P., and Giese, M.A. (2012). Neural theory for the perception of causal actions. Psychol. Res. *76*, 476–493.
7. Ullman, S., Harari, D., and Dorfman, N. (2012). From simple innate biases to complex visual concepts. Proc. Natl. Acad. Sci. USA *109*, 18215–18220.
8. di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. Exp. Brain Res. *91*, 176–180.
9. Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. Science *308*, 662–667.
10. Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J.K., Thier, P., Giese, M.A., and Casile, A. (2011). View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. Curr. Biol. *21*, 144–148.
11. Umiltà, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., and Rizzolatti, G. (2001). I know what you are doing. a neurophysiological study. Neuron *31*, 155–165.
12. Bonini, L., Maranesi, M., Livi, A., Fogassi, L., and Rizzolatti, G. (2014). Ventral premotor neurons encoding representations of action during self and others' inaction. Curr. Biol. *24*, 1611–1614.
13. Caggiano, V., Pomper, J.K., Fleischer, F., Fogassi, L., Giese, M., and Thier, P. (2013). Mirror neurons in monkey area F5 do not adapt to the observation of repeated actions. Nat. Commun. *4*, 1433.
14. Barraclough, N.E., Keith, R.H., Xiao, D., Oram, M.W., and Perrett, D.I. (2009). Visual adaptation to goal-directed hand actions. J. Cogn. Neurosci. *21*, 1806–1820.
15. Bonini, L., Serventi, F.U., Simone, L., Rozzi, S., Ferrari, P.F., and Fogassi, L. (2011). Grasping neurons of monkey parietal and premotor cortices encode action goals at distinct levels of abstraction during complex action sequences. J. Neurosci. *31*, 5876–5886.
16. Goodman, N.D., Ullman, T.D., and Tenenbaum, J.B. (2011). Learning a theory of causality. Psychol. Rev. *118*, 110–119.