Psychon Bull Rev (2017) 24:503–509
DOI 10.3758/s13423-016-1108-4

BRIEF REPORT

What's she doing in the kitchen? Context helps when actions are hard to recognize

Moritz F. Wurm[1,2,3] · **Ricarda I. Schubotz**[3,4]

Published online: 6 July 2016

Abstract Specific spatial environments are often indicative of where certain actions may take place: In kitchens we prepare food, and in bathrooms we engage in personal hygiene, but not vice versa. In action recognition, contextual cues may constrain an observer's expectations toward actions that are more strongly associated with a particular context than others. Such cues should become particularly helpful when the action itself is difficult to recognize. However, to date only easily identifiable actions were investigated, and the effects of context on recognition were rather interfering than facilitatory. To test whether context also facilitates action recognition, we measured recognition performance of hardly identifiable actions that took place in compatible, incompatible, and neutral contextual settings. Action information was degraded by pixelizing the area of the object manipulation while the room in which the action took place remained fully visible. We found significantly higher accuracy for actions that took place in compatible compared to incompatible and neutral settings, indicating facilitation. Additionally, action recognition was slower in incompatible settings than in compatible and neutral settings, indicating interference. Together, our findings demonstrate that contextual information is effectively exploited during action observation, in particular when visual information about the action itself is sparse. Differential effects on speed and accuracy suggest that contexts modulate action recognition at different levels of processing. Our findings emphasize the importance of contextual information in comprehensive, ecologically valid models of action recognition.

Keywords Action recognition · Scene · Semantic priming · Object

Electronic supplementary material The online version of this article (doi:10.3758/s13423-016-1108-4) contains supplementary material, which is available to authorized users.

✉ Moritz F. Wurm
moritz.f.wurm@gmail.com

[1] Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA

[2] Center for Mind/Brain Sciences, University of Trento, Rovereto TN, Italy

[3] Max Planck Institute for Neurological Research, Cologne, Germany

[4] Institute of Psychology, University of Münster, Münster, Germany

Spatial environments such as places and rooms (*contextual settings,* hereafter) are often indicative of certain classes of actions: In kitchens we prepare food, and in bathrooms we engage in personal hygiene, but not vice versa. Observers can exploit the statistical probability of the co-occurrence of actions and contextual settings to constrain their expectations toward actions that are more likely to take place in a particular setting than others. Crucially, actions unfold comparably slowly over time, whereas contextual settings are recognized with exposure durations below 100 ms (Biederman, Rabinowitz, Glass, & Stacy, 1974), and their semantic content is activated within 300 ms (Bar, 2004; Ganis & Kutas, 2003). Hence, preactivation of action information via associated contextual settings should be beneficial for the efficient processing of action information.

In object recognition, the facilitatory influence of contextual settings is well documented. For example, objects in compatible contextual settings are recognized faster (Boyce & Pollatsek, 1992) and more accurately (Barenholtz, 2013; Boyce, Pollatsek, & Rayner, 1989; Davenport & Potter, 2004; Palmer, 1975) as compared to objects in incompatible and semantically neutral settings. Current models suggest that

object recognition is constrained by top-down predictive signals via context-based associations, which facilitates recognition of the object in that particular context (Bar, 2004).

By contrast, evidence for contextual facilitation of action recognition is lacking. Instead, incompatible contextual settings have been demonstrated to interfere with action recognition: Participants need longer time to recognize actions in incompatible compared to compatible and neutral settings (Wurm & Schubotz, 2012). Moreover, neural activity in the left inferior frontal gyrus (IFG) increases when actions are perceived in incompatible versus compatible and neutral settings (Wurm & Schubotz, 2012). These effects were interpreted as increased effort in semantic integration of the observed action into an overarching action goal that is reconcilable with the incompatible setting. Although these findings clearly demonstrate selective associations between actions and contextual settings, it is unclear if these associations are effectively exploited in a facilitatory way.

Crucially, contextual information should become particularly beneficial for action recognition when actions are hard to identify. In support of this view, contextual effects on object recognition have been shown to dissociate between facilitation and interference as a function of general recognizability (Palmer, 1975). Notably, in Wurm and Schubotz (2012), actions were very easy to recognize, which leaves open the possibility that putative facilitatory effects were masked by ceiling effects. In this study we therefore investigated the effect of context on the recognition of actions that are hard to identify. In the case of object-directed actions, the most important (primary, hereafter) sources of information are fine and coarse movement kinematics of fingers, hands, and arms, and the objects involved in the action. In daily life, we often observe actions from some distance, or our view on an action is partly occluded (e.g., by other objects, persons, or the actress herself). Such factors will most likely affect the recognition of manipulated objects and fine motoric movements of the hand, whereas coarse postures and movements of the arms often remain recognizable. We hypothesized that in such typical situations, the influence of contextual information—activation of associated action semantics, possibly in addition to priming of object information—become particularly effective as reflected in improved recognition of the action.

Participants observed video clips of context-specific actions (e.g., hammering, cracking an egg). The recognizability of the actions was lowered by pixelizing the hands and the involved objects. As a result, hand postures, fine motoric finger movements, and objects could not be identified anymore whereas the coarse movement kinematics of the arms and hands remained intact. The actions took place in compatible, incompatible, and neutral settings (see Fig. 1a). The neutral setting consisted of a white background that did not bias the participants toward a specific contextual affiliation. Thereby, the neutral condition served as a baseline condition to dissociate facilitatory from interference effects. Participants were instructed to press a button as soon as they recognized the presented action. After the button press, they had to verbally name the action (see Fig. 1b). We collected the rate of correctly recognized actions and reaction times (RTs) to correct responses. Facilitatory effects were expected to manifest as higher recognition rates and faster responses to compatible as compared to incompatible and neutral settings, whereas interference effects were expected to manifest as lower recognition rates and slower responses to incompatible as compared to compatible and neutral settings.

Material and method

Participants

Forty-five healthy naïve volunteers (21–42 years, mean 26.4 ± 4.01 *SD*, 27 females, four left-handed) participated in the study. Any participant whose mean recognition rate was greater or less than two standard deviations from the group mean was excluded from further analyses ($N = 1$). All participants had normal or corrected-to-normal vision. Participants gave written consent before the experiment. Data were handled anonymously.

Stimuli

Thirty context-specific actions from the experiment reported in Wurm and Schubotz (2012) were used. Actions were object manipulations specific for the contextual settings *office*, *kitchen*, or *workshop*. Actions took place in compatible, incompatible, and neutral contextual settings (factor contextual compatibility). The contextual settings could be either compatible or incompatible to the action (e.g., a workshop action such as *hammering* was compatible with the context *workshop*, whereas the same action was incompatible with the contexts *kitchen* and *office*). Contextual settings were specified by the background, the working surface, and three to five context-specific stationary objects (e.g., computer screen, coffee machine, grinding machine). The neutral setting consisted of a uniform white surface without any corners.

In each context, actions were filmed from an identical allocentric perspective, providing a convenient view on both the object manipulation and the contextual setting (see Fig. 1a). At video onset, hands were positioned left and right to the to-be-manipulated objects with the palms on the working surface. The grasping of the objects followed about 680 ms after video onset. Each action was practiced 3 to 4 times before the filming of the actions in the three contexts, thus, the execution was basically identical across all the contexts. Each

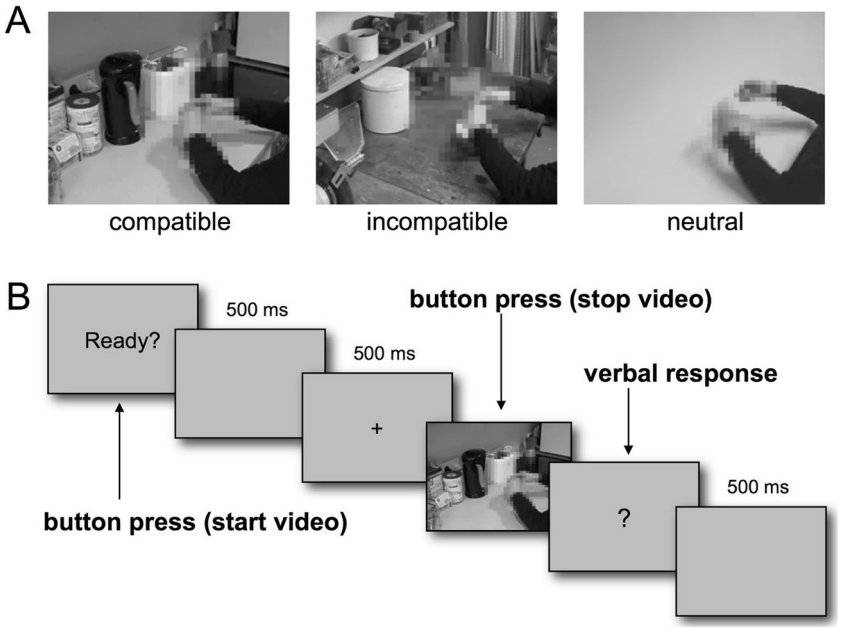

Fig. 1 (**a**) Example video frames of one action (cracking an egg) in the three experimental conditions. (**b**) Experimental trial design. Participants started the video presentation self-paced by button press. They were instructed to press the button again as soon as they recognized the action. After the button, press the video was replaced by a question mark indicating the participant to name the action. For details, see the Method section

video had a length of 3 s, a presentation rate of 25 frames per second, and a display width and height of 720 × 576 pixels.

We manipulated the recognizability of the actions by pixelizing the area of the video where the object manipulation was shown. The pixelized area was circular with a radius of 185 pixels. Pixelization consisted of averaging the gray values of pixels in 15 × 15 grid squares (see Fig. 1a). The pixelization resolution was chosen so that the objects were not identifiable in any of the static frames of the video (i.e., in the absence of movement information). After pixelization, objects could not be identified anymore whereas the action could still be inferred from movement kinematics (e.g., wrist rotations, horizontal and vertical arm trajectories; see Supplementary Data for video examples). A separate control experiment confirmed that objects were unrecognizable: Participants from a different sample ($N = 18$) were presented with static images of the first video frame (where objects were not occluded by the actress' hands) using the same procedure as in the main experiment. Participants were asked to name at least one of the objects in the pixelized area. Objects were identified correctly in 5.9% (±1.5 *SEM*) of the actions (see Supplemental Material for item-specific results).

Design and procedure

Design and procedure (see Fig. 1b) were identical to the behavioral experiment reported in Wurm and Schubotz (2012): Participants were seated approximately 60 cm away from a computer screen and next to the experimenter. Trials started self-paced by pressing a button with the right index finger, followed by a short fixation phase (500 ms blank screen, 500 ms fixation cross at the center of the screen). Videos appeared at the center of the screen, subtending approximately 13.6 × 10.5° of visual angle. Participants were instructed to press the button as soon as they recognized the action (i.e., during video presentation). If they did not recognize the action, no button press was required. After the video, a question mark appeared at the center of the screen indicating the participants were to name the action. After the button press during video presentation, the video stopped and was immediately replaced by the question mark. Participants had to name the recognized action using a single verb or short phrases using the infinite (e.g., *zeichnen* = *to draw*). Depending on the action, participants were required to name the objects involved (e.g., *sharpening pencil*) or not (e.g., *painting*). In case the participants did not recognize the action, they were asked to either guess or to answer with "I don't know." Each trial ended with a blank screen for 500 ms. The procedure was practiced using three actions that were different from those used in the experiment. Verbal responses were recorded using the presentation software (Presentation 13.1, Neurobehavioral Systems, Berkeley, CA). The experimenter indicated by button press that was not visible to the participant whether a verbal response was given and if the response was correct or not. For RTs, only correctly answered trials with a button press delivered during the video presentation entered the statistical analysis. Responses delivered after the video ended were treated as invalid because they were not directly linked to action information processing, and therefore not suited to analyze temporal aspects of the influence of context on action recognition.

As in Wurm and Schubotz (2012), each participant watched each of the 30 actions once during the experiment (10 trials per condition). Stimuli were balanced across participants so that

groups of three participants watched a complete set of the 90 stimuli (30 actions × 3 contexts). The occurrence of contexts was balanced within participants so that each participant saw each of the three contextual settings 6 to 7 times and the neutral setting 10 times. The trial order was counterbalanced so that transitions of settings (office, kitchen, workshop, neutral) and transitions of conditions (compatible, incompatible, neutral) occurred equally often; that is, for each participant (=30 trials) each of the 16 possible transitions between the four contextual settings occurred 1 to 2 times, and each of the nine possible transitions between the three conditions occurred 3 to 4 times.

Results

A repeated-measures ANOVA revealed a significant effect of the three-level factor contextual compatibility on the rate of correctly identified actions ($F_{(2,\ 86)} = 4.79$, $p = 0.011$, $\eta^2 = 0.07$). Participants recognized the actions more often when presented in a compatible contextual setting (mean ± standard error of mean, 0.54 ± 0.03) compared to incompatible (0.44 ± 0.02; one-tailed paired t test, $t_{(43)} = 2.74$, $p = 0.004$) or neutral settings (0.43 ± 0.03; one-tailed paired t test, $t_{(43)} = 2.2$, $p = 0.002$). The recognition rate did not differ significantly between actions in neutral and incompatible settings (one-tailed paired t test, $t_{(43)} = 0.32$, $p = 0.75$; see Fig. 2a). The difference between effects of compatibility (neutral–compatible) and incompatibility (incompatible–neutral) on error rates was significant (two-tailed paired t test, $t_{(43)} = 2.7$, $p = 0.009$) which demonstrates the specificity of the compatibility effect on recognition rates.

Regarding RTs, we found a significant effect of contextual compatibility ($F_{(2,\ 54)} = 3.49$, $p = 0.037$, $\eta^2 = 0.06$; see Fig. 2b): actions were recognized slower when they took place in incompatible contextual settings (mean ± standard error of mean, 2,568 ± 43 ms) compared to compatible (2,477 ± 40; one-tailed paired t test, $t_{(29)} = 1.49$, $p = 0.073$; difference 91 ms) or neutral contextual settings (2,444 ± 52; one-tailed paired t test, $t_{(29)} = 2.69$, $p = 0.006$; difference 124 ms). RTs did not differ significantly between the neutral and the compatible condition (one-tailed paired t test, $t_{(29)} = 0.11$, $p = 0.9$). The difference between effects of compatibility (neutral–compatible) and incompatibility (incompatible–neutral) on RTs was significant (two-tailed paired t test, $t_{(29)} = 2.1$, $p = 0.046$), which demonstrates the specificity of the incompatibility effect on RTs. Note that the analysis of RTs relied on fewer trials because only button presses of correct responses delivered before video end were treated as valid (18% of all trials; correct responses after video end: 27%, false responses: 55%). Therefore, only 28 of 44 participants provided valid trials for all three conditions and thus entered the ANOVA. Similarly, 30 of 44 participants provided valid trials for at least two conditions and thus entered the t tests. The mean percentage of valid trials per participant was 23% (compatible), 30% (incompatible), and 20% (neutral). This reduction might have lowered the power, which could explain why the RT effects were stronger in Wurm and Schubotz (2012).

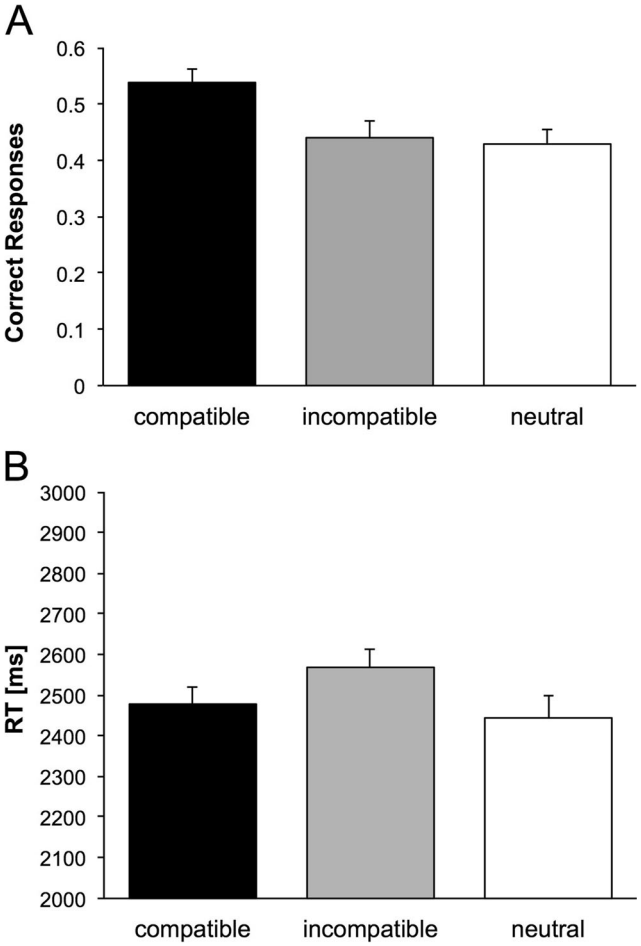

Fig. 2 (**a**) Effects of contextual compatibility on recognition rates. Actions were significantly better recognized when they took place in compatible settings. (**b**) Effects of contextual compatibility on reaction times (RT). Action recognition was significantly slower when actions took place in incompatible settings. Error bars indicate standard error of mean

After the experiment, participants were asked whether they noticed anything odd in the videos. Only 23% (10 out of 43 participants) spontaneously reported that some actions were performed in action-incompatible settings. In a second step, we explicitly asked whether the actions always took place in their typical rooms. Here, 65% (28 out of 43 participants) reported that some actions were performed in action-incompatible settings. ANOVAs for RTs and recognition rates using the ratings as between-subjects factors revealed no interactions with contextual compatibility (all ps > 0.8).

Discussion

This study tested the hypothesis that contextual information becomes relevant for action recognition when primary sources

of information that are typically exploited during action observation (i.e., the manipulated object and fine motor kinematics) are obstructed in the stimulus. To this end, we analyzed the influence of contextual settings on the recognition of actions that were pixelized and thus more difficult to identify. Recognition rates were selectively enhanced in compatible settings (reflecting facilitatory effects), whereas RTs were selectively prolonged in incompatible settings (reflecting interference effects). Our results extend previous findings (Wurm & Schubotz, 2012) by demonstrating that contextual settings can enhance (and not only disturb) action recognition. Crucially, by pixelizing objects, we show that the influence of context on action recognition is mediated in absence of object information. Taken together, by providing evidence for effects of context–action associations we complement studies that demonstrate effects of context–object (Davenport & Potter, 2004; Palmer, 1975) and object–action (Bach, Nicholson, & Hudson, 2014; Schubotz, Wurm, Wittmann, & von Cramon, 2014; Thioux & Keysers, 2015) associations.

In the following, we discuss facilitatory and interference context–action effects and conclude with a discussion of the idea that the dissociation of these effects points to context–action interactions at different levels of processing.

Compatible contextual settings enhance action recognition

Actions were recognized with about 25% higher accuracy when they took place in compatible as compared to incompatible or neutral settings. This novel effect can be interpreted as facilitation via preactivation of action knowledge by the setting: Following the principles of statistical Hebbian learning (Hebb, 1949; Munakata & Pfaffly, 2004), co-occurrence of actions in their typical settings should lead to enhanced associative strengths between actions and settings. Activation of the setting therefore enhances the excitability of strongly associated action information to a higher degree than weakly associated action information. Thus, context-selective preactivation of action information increases the likelihood to activate the to-be-identified action. Such preactivation should be particularly useful when the action is not easy to identify because primary action information (i.e., information that is directly relevant for action recognition, such as object and manipulation information) is sparse. Indeed, our experimental manipulation of impeding recognition of the object and fine motoric finger movements, and thereby muting primary channels to action recognition, resembles a common case of our daily life: For example, actions are often partially hidden by other objects or persons, or they are too far away to be unambiguously identified. Our findings suggest that in such cases, contextual settings provide additional sources of information that have the capacity to narrow down the search space toward expectable actions. Future studies should further examine the effects of context in more naturalistic conditions. In addition to constraining action expectations, depletion of object information could enhance the perceptual analysis of manipulation information.

Facilitatory effects were reflected in higher recognition rates, but not in higher speed of recognition. Contextual setting information is already available after at least 300 ms (Bar, 2004; Biederman et al., 1974; Ganis & Kutas, 2003), whereas actions unfold over time. Excitation of conceptual action information via contextual settings should therefore occur earlier than excitation via movement kinematics of the action. Hence, preactivation of action information should result in faster access of the observed target action. We cannot exclude that compatibility effects were present but remained undetected because of faster processing time of the neutral setting, which is less rich compared to the compatible and incompatible setting. As discussed in detail in Wurm and Schubotz (2012), neutral settings cannot unambiguously dissociate facilitation from interference effects on RTs and therefore need to be interpreted with caution. Note, however, that different processing speeds of neutral versus compatible and incompatible settings should be unlikely to have substantial effects on recognition accuracies discussed in the previous section.

Incompatible contextual settings slow down action recognition

Responses to actions taking place in incompatible settings were delivered about 100 ms later relative to compatible or neutral settings. This effect size is in a similar range as the effect size found for RTs to the recognition of natural action in incompatible settings (Wurm & Schubotz, 2012). Interestingly, when explicitly asked, 35% of participants did not notice any incompatibility between actions and contextual settings.

Inhibitory effects of incompatible contexts on action recognition can be explained in different, not mutually exclusive, ways: First, observation of incompatible settings might activate a set of actions that does not include, and thereby distract from, the target action. Hence, higher RTs might reflect longer search for the target action. However, in this case one should also expect shorter RTs for compatible contexts, which were not observed (but see our discussion in the previous section). Second, delayed responses to actions in incompatible settings might be due to a conflict that occurs *after* recognition of the action. Thus, once the action is identified, incompatibility with the setting might interfere with the response. Because the pixelized actions were generally hard to identify, incompatibility might trigger a reanalysis of the action, which in turn would delay the response. Moreover, incompatibility might cause a conflict at a higher level of interpretation because the recognized action is difficult to reconcile with the incompatible setting. Accordingly, increased RTs could reflect an

attempt to integrate the observed action into an overarching action that is in agreement with the setting. To this end, alternative or even implausible explanations might be constructed that are not in the common repertoire of a contextual settings' action scripts (Schank & Abelson, 1977). This interpretation is supported by neuroimaging evidence: The inferior frontal gyrus (IFG), a brain region known to be involved in semantic integration (Badre & Wagner, 2007), is more strongly activated when observed actions take place in incompatible compared to compatible settings (Wurm & Schubotz, 2012). Likewise, the IFG is modulated by the ease of integrating separate action steps into a common overarching goal (Hrkać, Wurm, & Schubotz, 2014; Wurm, Hrkać, Morikawa, & Schubotz, 2014), in line with our proposal that increased RTs reflect interference at the level of semantic integration.

Contexts, objects, and manipulations are integrated at different levels of action processing

Contextual settings, objects, and manipulations/movements are processed by distinct but anatomically connected neural substrates (Epstein, 2005; Grill-Spector, Kourtzi, & Kanwisher, 2001; Kravitz, Saleem, Baker, & Mishkin, 2011; Schubotz & von Cramon, 2009; Schubotz et al., 2014; Wurm & Schubotz, 2012). Associative strengths between contextual settings, objects, and manipulations may be functions of statistical co-occurrence (Bar, 2007) building tightly interconnected semantic networks, recently coined context-object-manipulation (COM) triads (Wurm et al., 2012). During action recognition, these three sources of information may affect each other following Bayesian principles: For each source, the likelihood is estimated based on the probability of co-occurrence with the two other sources. From a neural perspective, probabilities of co-occurrence are expressed in the association strengths between the three sources. During perception of an action scene (e.g., observing someone cracking an egg in a kitchen), the neural representations of the context (*kitchen*), object (*egg*), and manipulation (*crushing*) may activate one another as a function of their association strengths (see Fig. 3). Consequently, if one source is absent in the stimulus, it can be predicted from the two remaining channels (Bayesian inference; Friston, 2005). In addition, based on these lateral "between-sources" constraints, the higher level action goal (*cracking an egg*) is activated in a bottom-up manner. Notably, contextual settings might preactivate not only objects and manipulations but also the action goal itself, as well as script knowledge (i.e., overarching long-term goals that are typically achieved in the specific setting, e.g., *making pancake*; see Fig. 3). Associations between the distinct levels may in turn preactivate action-relevant information at lower levels of the hierarchy in a top-down manner. For example, with regard to this experiment, top-down predictions about perceptual object information could be matched with the pixelized object shape, which, in case of accurate prediction, produces a low prediction error. Taken together, different sources of action-relevant information serve as priors that together estimate the most likely action by reducing prediction error at all levels of the action processing stream (Kilner, Friston, & Frith, 2007). A yet unsolved issue is whether COM associations develop predominantly on a perceptual "token" level (associations between representations of

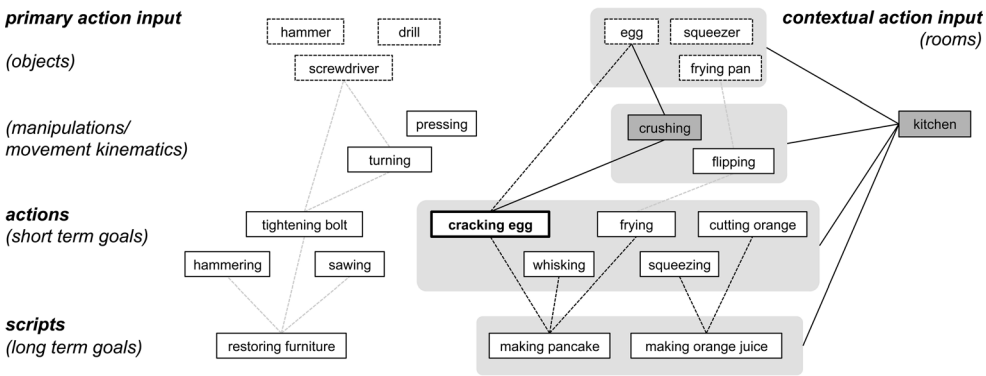

Fig. 3 Simplified schematic illustration of the possible associations between primary and contextual action input representations (upper two rows), action representations (third row), and higher level script knowledge (lower row). Dark gray text boxes indicate representations directly activated by the stimulus in the experiment (context and movement kinematics), light gray areas indicate indirectly activated representations. Dotted text boxes indicate object representations that were not recognizable in the stimulus. Solid lines indicate associations from directly activated input representations, black dotted lines indicate indirectly activated bottom-up/top-down associations between levels of the action hierarchy, gray dotted lines indicate not activated associations. For clarity, not all possible associations are depicted. The bold text box indicates the action with the highest sum of activation. Note that the descriptions of manipulations/movement kinematics are simplified and refer to more complex movement patterns (e.g., *crushing* is intended to describe the movement of crushing an egg shell with the thumbs to let the egg yolk flow out, *flipping* is intended to describe the movement to flip a pancake with a pan). Note also that some manipulations are expected to be less context specific (e.g., *turning*—i.e., a wrist-elbow rotation—could be associated with tightening a bolt but also with, e.g., sharpening a pencil; cf. Watson & Buxbaum, 2014)

concrete object exemplars, object-specific movement kinematics, and specific contextual settings) or on a conceptual "type" level that is independent of concrete instantiations of contexts, objects, and manipulations. Because it is more plausible that statistical regularities in action scenes are based on experience from multiple, typically highly variant situations, we assume that COM triadic relationships are stronger between perceptually invariant representations. Note, however, that boundaries between perceptual and conceptual levels are fuzzy, and therefore the COM triad model is inevitably simplistic in this respect.

In addition, other sources, such as information derived from the actor (Hrkać et al., 2014; Wurm, von Cramon, & Schubotz, 2011) or from previous events (Hrkać, Wurm, Kühn, & Schubotz, 2015), are likely to serve as action priors as well and to shape action recognition in a similar way.

Conclusions

Our study shows that action-compatible contexts enhance action recognition rates, whereas action-incompatible contexts prolong action recognition times. We thereby demonstrate that not only context–object and object–action but also context–action associations are effectively exploited in visual perception. We suggest that contextual settings interact with action recognition at different levels of processing. The identification of further contextual action-relevant sources of information, such as actor identity and episodic memory about previous events, and the isolation of the different interactions between these sources should be subject to future research to establish a comprehensive, ecologically valid model of the multidimensional process of action perception.

Acknowledgments We would like to thank Olivia Cheung for helpful comments on the manuscript, and Nadiya El-Sourani and Marcin Lipski for assistance in data acquisition.

References

Bach, P., Nicholson, T., & Hudson, M. (2014). The affordance-matching hypothesis: How objects guide action understanding and prediction. *Frontiers in Human Neuroscience, 8,* 254.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia, 45*(13), 2883–2901.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617–629.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Science, 11*(7), 280–289.

Barenholtz, E. (2013). Quantifying the role of context in visual object recognition. *Visual Cognition, 22*(1), 30–56.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W., Jr. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology, 103*(3), 597–600.

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psycholboy: Learning, Memory, and Cognition, 18*(3), 531–543.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 556–566.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*(8), 559–564.

Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition, 12*(6), 954–978.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 360*(1456), 815–836.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Brain Research. Cognitive Brain Research, 16*(2), 123–144.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*(10-11), 1409–1422.

Hebb, D. (1949). *The organisation of behaviour*. New York, NY: Wiley.

Hrkać, M., Wurm, M. F., & Schubotz, R. I. (2014). Action observers implicitly expect actors to act goal-coherently, even if they do not: An fMRI study. *Human Brain Mapping, 35*(5), 2178–2190.

Hrkać, M., Wurm, M. F., Kühn, A. B., & Schubotz, R. I. (2015). Objects mediate goal integration in ventrolateral prefrontal cortex during action observation. *PLoS ONE, 10*(7), e0134316.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing, 8*(3), 159–166.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience, 12*(4), 217–230.

Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science, 7*(2), 141–148.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3,* 519–526.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.

Schubotz, R. I., & von Cramon, D. Y. (2009). The case of pretense: Observing actions and inferring goals. *Journal of Cognitive Neuroscience, 21*(4), 642–653.

Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in Psychology, 5,* 636.

Thioux, M., & Keysers, C. (2015). Object visibility alters the relative contribution of ventral visual stream and mirror neuron system to goal anticipation during action observation. *NeuroImage, 105,* 380–394.

Watson, C. E., & Buxbaum, L. J. (2014). Uncovering the architecture of action semantics. *Journal of Experimental Psychology: Human Perception and Performance, 40*(5), 1832–1848.

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: Contextual information modulates action recognition. *NeuroImage, 59*(2), 1551–1559.

Wurm, M. F., von Cramon, D. Y., & Schubotz, R. I. (2011). Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. *Human Brain Mapping, 32*(12), 2141–2150.

Wurm M. F., Cramon D. Y., Schubotz R. I. (2012) The context-object-manipulation triad: cross talk during action perception revealed by fMRI. *Journal of Cognitive Neuroscience 24*(7), 1548–1559.

Wurm, M. F., Hrkać, M., Morikawa, Y., & Schubotz, R. I. (2014). Predicting goals in action episodes attenuates BOLD response in inferior frontal and occipitotemporal cortex. *Behavioural Brain Research, 274,* 108–117.