Radboud University Radboudumc

DONDERS INSTITUTE

Action Recognition in Context

An fMRI Multivoxel Pattern Analysis Paradigm with Motion-Captured Actions

written by

Anna B.C. Trimborn

Thesis submitted for the degree of

Master of Science
Cognitive Neuroscience

Radboud University Nijmegen, The Netherlands

Author: Anna B.C. Trimborn
Student Number 1013433

Dates: Submission: July 1st 2020, Defense: July 8th 2020

Thesis Advisor: Dr Sebo Uithol

2nd Reader: Prof Harold Bekkering

Abstract

As humans, our everyday life is inherently social. To interact in a world shaped by social interaction, it is crucial to recognise and understand other agent's actions. While action recognition is widely studied in psychology and cognitive neuroscience, contextual information, for instance the scene an action is embedded in, is often left out. There is promising evidence underlining the importance of context and its congruency on action recognition. The neuronal correlates of this context influence on action recognition are, however, not clear yet. To test action recognition in context, we designed a novel task in which first an indoor scence context image alone is presented, and then motion-captured action videos are superimposed on the image followed by a forced-choice task. We tested the new paradigm first behaviourally (N=25), and then used functional Magnetic Resonance Imaging (fMRI) to investigate brain activation while participants performed the task. The behavioural results show a context-congruency effects in recognition accuracy, with significant effects of incongruent-congruent and incongruent-neutral context settings on action recognition. For the fMRI part we at present have a sample of N=8 participants, which makes the statistical tests intended impossible. Design, analysis pipeline and expectations of this fMRI paradigm with decoding techniques are still discussed in the present thesis.

Note

All code used to acquire the data in the behavioural and fMRI experiments as well as the stimuli dataset is available online on a publically available repository (please consider the **GitLab repository** of this study for further visualisations of the motion-captured actions and scripts used).
While I could have attached source code in the appendix, a more compact and ready-to-reproduce version is uploaded in the repository. The behavioural data I used for the analysis is, of course, anonymised.

Contents

1 Introduction

Our everyday life is inherently shaped by social interactions ranging from buying groceries to more complex ones such as working in teams or being in relationships. Humans are uniquely specialized to interact with each other to survive (Gallese et al., 2004). For those social decisions as well as to make predictions regarding one's own behaviour, it is crucial to recognise and understand other agent's actions (Frith and Frith, 1999).

The manifold of cognitive processes associated to action understanding is typically grouped under umbrella terms such as 'Theory of Mind' (Saxe and Baron-Cohen, 2006) 'Mindreading' (Heyes and Frith, 2014), 'mirroring'(Borroni et al., 2011) and 'mentalizing' (Ondobaka et al., 2013). A lack of the abilities to understand actions or intentions of social interaction-partners can affect one's every-day life and personal relations massively. Such issues are an essential aspect of disorders such as autism spectrum disorder (Chambon et al., 2017), or depression (Bora and Berk, 2016) and covary with multiple other frequent mental disorders (Kim et al., 2011).

One of the processes associated with living and interacting in social contexts is recognising other individuals' actions based on visual input. Decades ago, mirrorneurons (Rizzolatti and Fabbri-Destro, 2010, Ferrari et al., 2003, Kilner et al., 2009) became a prominent topic in neuroscience, and with them the finding that the human motor system becomes activated during action observation (a process called *motor resonance* Uithol et al., 2011). Related theories often take this activation as necessity to recognise observed actions and infer the *goals and intentions* of actions. The usage and explanatory power of an experimental study with regards to those concepts (such as action recognition, understanding or 'intention') often remains unclear (for an elaborate overview see: Uithol et al., 2011).

Theoretically speaking (i.e. from a formal modelling perspective), action recognition is a computationally expensive process with a high degree of complexity (Zeppi and Blokpoel, 2017, Blokpoel et al., 2013). In the attempt to make their research question tractable, experimental researchers often reduce their stimuli to the most isolated form to eliminate confounding factors (overview of related issues see Matusz et al., 2018). Hence, context information is often cut out of the experimental design (for instance, Cattaneo et al., 2012).

Scene information, however, can have a crucial impact on the recognition of action (Wokke et al., 2016). When trying to decipher if a *person is driving a screw or sharpening a pencil*, not only the kinematic information about the actual movement

(position of the body over time) is taken into account. Contextual cues - such as the setting in which an action is performed in - have been shown to influence the accuracy of action recognition (Amoruso et al., 2016; Wurm and Schubotz, 2012).

Behaviourally, it has been shown that context information (e.g. indoor scene images) facilitates action recognition when congruent with the action presented (looking at reaction times Wurm and Schubotz, 2012) and impedes recognition when incongruent (measured with reaction times and accuracies Wurm and Schubotz, 2017). In these studies congruent contexts meant performing an action in the location it is most commonly associated with, such as brushing one's teeth in the bathroom. Incongruent contexts on the other hand are randomly chosen locations that do not match the action association mentioned above, e.g brushing one's teeth in the kitchen. The control condition most commonly included the performance of the action in an empty room.

Investigating the neuronal correlates, Iacoboni et al. (2005) used a paradigm in which similar actions (grasping a cup) would result in different associated intentions depending depending on two different contexts presented. Here, grasping a cup to clean up vs. grasping a cup to drink. The authors report significantly more activation in the inferior frontal gyrus and the ventral premotor cortex (PMv) for action presented in congruent context, compared to others. Because the PMv has previously been associated with hand movements, the authors conclude that mirror-neuron areas in the motor system are not only involved in the recognition, but also the understanding of the goal and intention of an action (Iacoboni et al., 2005). An important caveat of the study, however, is that the action sequences were slightly altered with the context manipulation, which is a confounding variable that could have biased context effects significantly.

Wurm & Schubotz (2012) tested different everyday actions in either congruent, incongrent and neutral with a univariate functional Magnet Resonance Imaging (fMRI) study. The authors showed brain regions in the ventrolateral prefrontal cortex with increased activation for incongruent context-action-presentations. The authors conclude that contextual information is taken into account at a high level of processing, even if the context is irrelevant to the given task (Wurm and Schubotz, 2012). In ecologically-relevant paradigms, however, context-information generally is task-relevant. Further, with the region-of-interest approach contrasting informative with less complex, empty rooms, it is questionable whether context-action-interactions were studied.

Using transcranial magnetic stimulation (TMS) to investigate effects of context on motor resonance, Amoruso et. al (2016) propose context effects to be happening at different time points of the action recognition process, suggesting facilitation

to happen early on and inhibition due to incongruency later in the process. The authors presented action videos in incongruent, congruent or 'ambiguous' context and then measured TMS motor evoked potentials (MEPs) to disentangle the time course (Amoruso et al., 2016). A drawback here are the constraints of TMS, importantly that the stimulation location is difficult to pinpoint exactly and is affecting the occurrence and size of the motor evoked potentials, which could bias the results, as those are the main dependent variable. Further, the study is not explanatory with regards to the particular neuronal correlates of the process.

These studies provide promising evidence for the importance of context, but the concrete involvement of context in action recognition on the level of brain activation is not clear yet. Hence, for the present study, we were interested in the following research questions: As the overarching research question we investigated, if available context information impacts action recognition. More specifically we were interested in the brain regions involved in processing context information and the brain regions that are primarily involved in action recognition, respectively. Finally, we aimed at answering the question, to which extent the areas involved in action recognition also involved in the processing of context information.

Establishing a causal context-effect is hard, because the brain activation when recognising motion in scenes is widespread and variable, so proving the exact involvement of a specific brain region is difficult. In our paradigm, we therefore investigate context involvement in action recognition with multivariate-pattern analysis. To this end, we designed a novel paradigm in which we investigated how actions are recognised under the presence of scene information, without the necessity of relying on imprecise theoretical constructs. In a tailored task, participants viewed point-light display videos of everyday actions (e.g. 'hammering' or 'chopping'), which are superimposed on context images (e.g. of a 'garage' or a 'kitchen'). Per trial, the first context image only was presented on the screen and then an action sequence was superimposed on top of the image. The context image was either congruent, incongruent or neutral regarding the simultaneously presented action. In a computer-based study we first tested context-involvement on a behavioural level. Based on Wurm et. al (2012) we hypothesised that a congruency effect is revealed by the accuracy data with the developed paradigm. Building on the behavioural results, this study was intended and designed to be tested with a full sample using functional Magnet Resonance Imaging (fMRI) to investigate the neuronal correlates of context effects. In an event-related within-subject design suitable for the analysis with Multivoxel Pattern Analysis (MVPA), we aimed at comparing patterns of evoked brain activity when participants perceived an action with context information (Poldrack, 2011, Görgen et al., 2018, Kriegeskorte and Kievit, 2013).

For the fMRI part we at present have a sample N=8 participants, which make the statistical tests intended impossible. Design, analysis pipeline and expectations are still discussed in the present thesis.

2 Methods

To test action recognition, we designed a novel task in which first the context alone is presented, and then action videos are superimposed on the image with a subsequent forced-choice task. We tested the new paradigm first behaviourally, and then used functional Magnetic Resonance Imaging (fMRI) to investigate brain activation during the performance of the task. The behavioural task and fMRI task are largely similar, with a few changes necessary to adjust the paradigm to the MRI-scanner environment and data requirements. Experimental design and procedure for both stages of the project are explained below.

2.1 Behavioural study

2.1.1 Participants

In the behavioural experiment, 25 participants were tested. One participant had to be excluded due to failure to adhere to instructions. Out of the remaining 24, 17 were female, and the age range was between 18 and 25 (with the mean age of 22.24 and a standard deviation of 4.26). All participants had normal or corrected-to-normal vision and reported no history of neurological or psychiatric disorders. The experiment was performed in Dutch, the participants were Dutch native speakers and were recruited through the Radboud University online recruitment system (Sona). They received €10,- for participation in the behavioural study (1.0 hour). Participants were informed about the study, data security and handling as well as their rights to withdraw at any time prior to their participation and gave their informed consent. The study was approved by the local ethics committee and complied with the declaration of Helsinki.

2.1.2 Experiment Design

To evaluate our motion-captured actions and check if our developed paradigm is suitable for answering our research questions, a behavioural experiment was conducted.

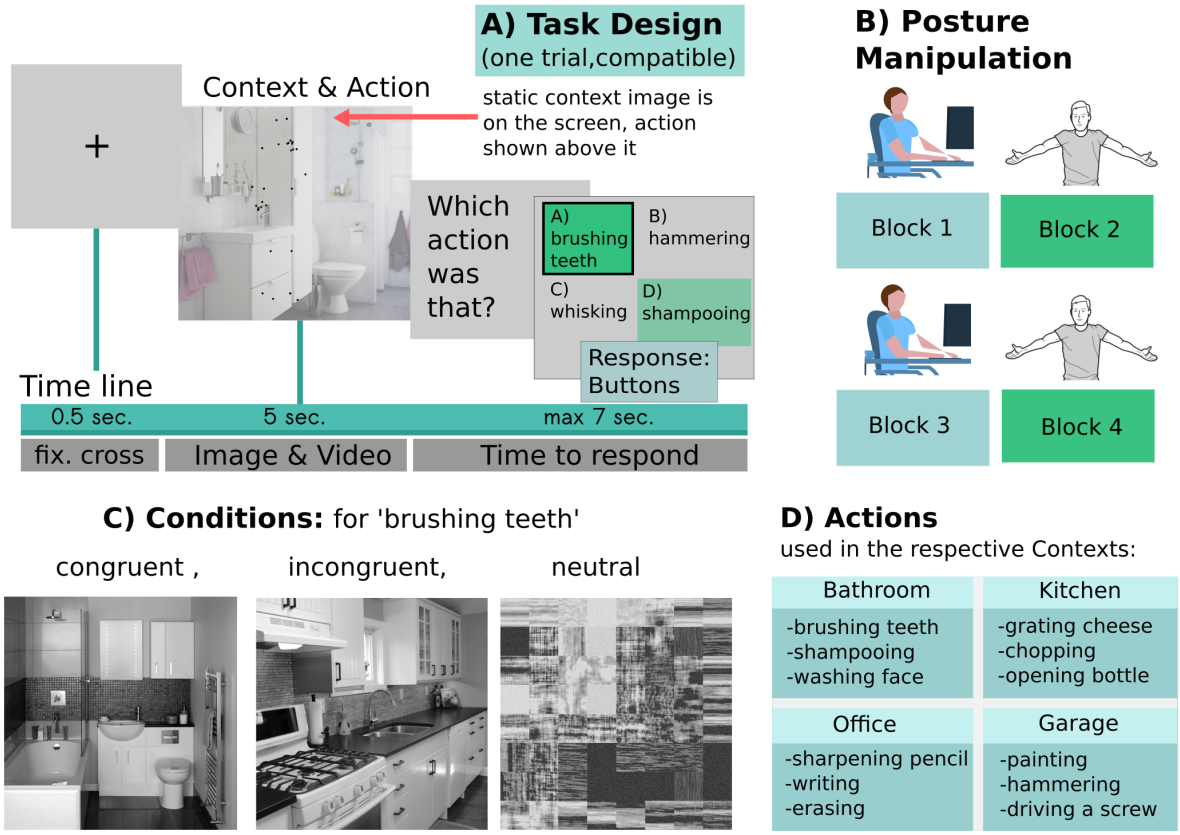


Figure 1: Behavioural Task Design

In a similar design as Wurm and Schubotz, 2012, we tested one action in different contexts. Every trial started with a fixation cross, subsequently (dependent on the condition), either a congruent, incongruent or neutral context image was shown (for further explanation on the stimuli see 1). An action video sequence was superimposed on this static context image. The action thereby was composed of single frames. The frame rate was recorded with 100 Hz, but then down-sampled to 25 Hz, equaling 300 frames per action. By adding a noise-jitter, meaning additional moving dots in the same size and colour as the point-light display action, we were able to make the actions harder to recognise, hence manipulating the task difficulty. When a sequence finished, the participant was faced with a multiple choice question with four answer alternatives corresponding to four buttons on the participant's keyboard.

The choice options were not fully random, one option was false but sampled from the same context as the correct response. The other two options were also false, but were sampled from different contexts. All 12 actions were presented twice in three conditions leading to a final number of 76 presentations per block. One trial took approx. 10 sec, which summed up to 60 min for the behavioural experiment. An example of one trial used in the behavioural experiment 'washing face' can be seen on the repository (**click here**). As De Lange et al. (2006) reported that behaviourally manipulating the posture match between observer and observed action agent has an influence on ac-

tion recognition, we introduced a similar posture manipulation, to get insights about underlying neuronal correlates already at the behavioural stage (de Lange et al., 2006).

2.1.3 Experimental Procedure

Upon arrival at the laboratory, participants received information about the study and when agreed, signed the informed consent. They were then seated on a desk, approximately 70cm away from a screen (screen specifications, 60hz). Subsequently, the subjects performed a training of the task, after which questions could be clarified. This is particularly important for the posture manipulation, which is not directly intuitive. Subsequently, participants saw every action sequence once in full and labelled with the corresponding word. Then the main experiment started, which contained four blocks, divided by short breaks facilitating participants concentration and instructions concerning the posture for the upcoming block.

For the posture manipulation, in two of the four blocks participants grasped two joysticks placed on the desk left and right respectively next to the computer screen while watching the action videos. Hence, their arms were spread roughly 80° degrees relative to the viewing direction. Consequently, the posture in those blocks did not match the upper limb actions, which presented in the action videos. In the other two blocks, participants arms were rested on the table in front of their body, more closely matching the posture of the figure in the action sequences. The start block was counterbalanced in an inter-subject way and then alternated per participant in a within-subject way.

After completion of the experimental task, all participants rated the actions in different contexts, with the question 'How likely is it that you XXX in this room' (in Dutch 'Hoe waarschijnlijk is het dat je in deze ruimte XXX?') to assess the typicality of a given action for each context.

2.1.4 Stimuli

2.1.4.1 Actions: Motion Capture

motion-capture recording Similar to previous studies using motion captured actions (Vanrie and Verfaillie, 2004), we recorded 16 everyday actions that are typically executed in one of four different contexts: kitchen, bathroom, garage and office. These everyday actions were all indoor, non-sports activities in which the upper body was most informative to recognise the action. We kept the whole-body figure, particularly

because we altered the size of the displays. Short action sequences of roughly 5s. were performed, to have a each action executed within a comparable time frame. We used half-profile variations with a common angle for each action as it was established that viewpoint is influencing action recognition performance (Jokisch et al., 2006). Additionally, we recorded four actions that were later used only in training trials and were not further taken into account in analyses. All actions were recorded with one actor, in one session, at a marked position in the room, resulting in a common viewpoint.

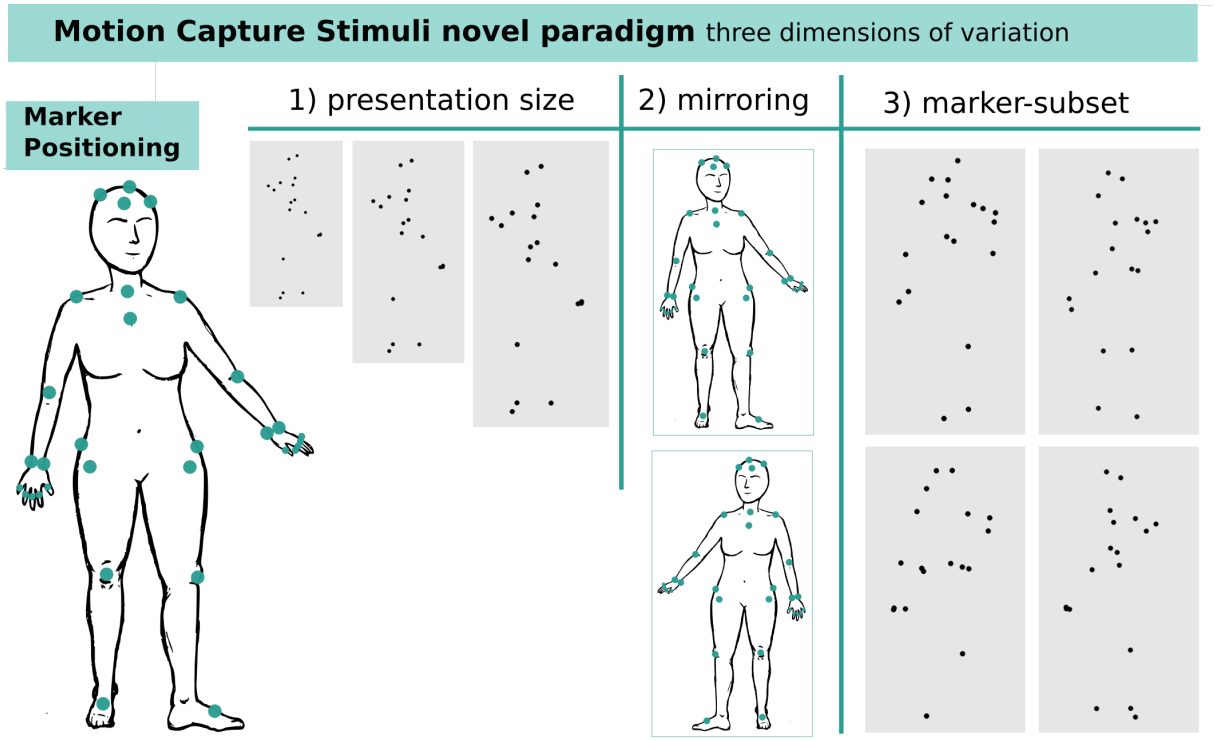


Figure 2: Positioning of Motion-Capture markers. **On the left**: full set of recorded markers visualised on a body schema. **On the right**: Three dimensions on which presentation of actions were altered, the combination of the three modifications yields 3 sizes $*$ 2 inversion variants $*$ 4 marker subsets $=$ 24 variants per action.

We used a 'Qualisys motion tracking system' comprised by 6 infrared cameras at the corners of a small room, which enabled us to capture 3D time series data of movements with 100 frames per second (which for the stimuli was down-sampled to 25 frames per sec). We used 32 Passive markers, that were registered by the cameras by reflecting the emitted light. The markers were carefully placed on selected key locations important to recognise the body. We based the initial set of marker locations on a previous study (by Jokisch et al., 2006), but then extended the set by 12 additional markers on the hands and wrists to get more detailed resolution in the upper body and limb movements. The markers were placed on joints, ankles and knuckles, as well as the sternum and hips. A headband with four markers attached was used to register the head movement (for the specific positioning on the agent's body: 2).

2.1.4.2 Context: Background Images

To manipulate context information in each trial, we presented images of the mentioned indoor scenes (kitchen, garage, bathroom, office) behind the point-light display videos. Indoor scenes differ highly in terms of their spectral properties and colour histograms. Those low-level visual features are shown to influence action recognition (Quattoni and Torralba, 2009 Watson et al., 2017). Particularly when training and testing a machine-learning classifier on neuronal activation, all low-level features of presented stimuli can distort how the classifier distinguishes conditions (Todd et al., 2013), which would be a heavily confounding factor. Hence, we transformed the background images into grey scale, and matched the spectral histograms of each image with a reference-histogram (see appendix) employing the **shine toolbox** (Willenbockel et al., 2010) for MATLAB (Mathworks, Natick Massachusetts, USA) with its *histMatch* function. Five images per context were shown, avoiding that the patterns of brain activity only represent retinotopic information.

For the 'neutral' backgrounds we went a step further and wrote a custom built script to generate neutral images from the whole set of used context images of the congruent and incongruent conditions. We first block and then subsequently fourier scrambled neutral images from the meaningful images to keep the complexity but distort any meaningful, visible objects (similar to Watson et al., 2017). The results for those neutral background images keep some local properties as well as global features all matched to the same reference histogram as the other images used.

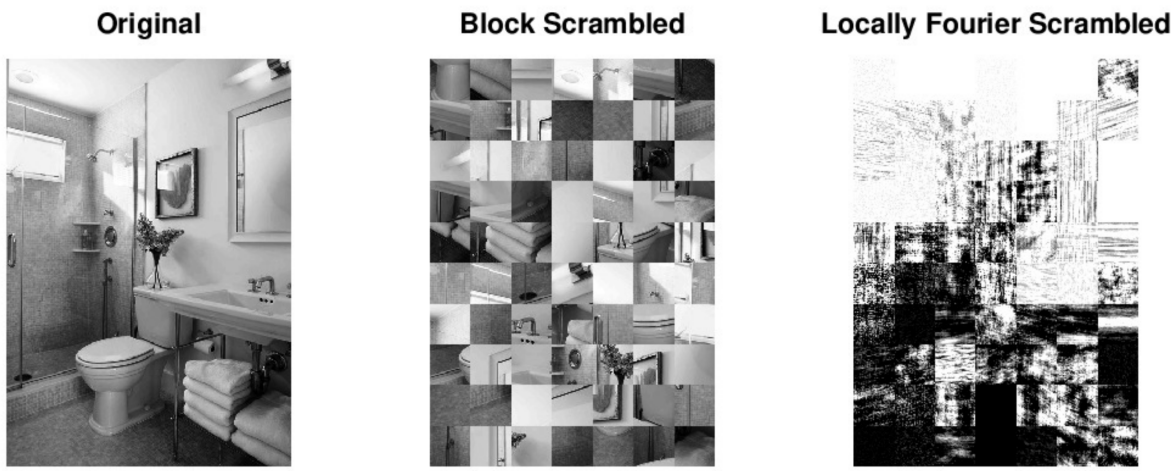


Figure 3: Different possible ways of scrambling for neutral images. We chose a combination of both, for the script see GitLab

2.1.5 Data Analysis

The response data from one participant had to be excluded from further analyses. This participant did not adhere to the experimental instructions of the posture manipulation, not even after repeated additional instructions.

For the analysis of the behavioural experiment, we used common frequentist statistical analyses using hypothesis testing with JASP (JASP Team (2020). JASP (Version 0.12.2)[Computer software]). The data of the dependent variable 'accuracy' was used in a full within-subject design in the comparisons, so repeated measure Analysis of Variance (rmANOVA) and post-hoc t-tests were conducted. Response time was not an essential part of the analyses. First, we plotted the difficulty of a given action through recognition responses (accuracy) for all actions averaged over all blocks and participants in the neutral condition, congruent and incongruent conditions.

To test our main research question, we subjected the dependent variable action accuracy to a two-way repeated measures ANOVA with the within-subjects factors of context congruency (congruent, incongruent neutral) and posture (wide, narrow). This analysis allows us to examine whether there is a context-congruency-effect on action recognition, whether posture had an effect on recognition performance, and a potential interaction between two factors. To check which conditions differed from each other, we conducted paired-sample post-hoc t-tests.

As a control measure, it was important to test whether participants performed significantly better in the later blocks of the experiment than in the first ones which could represent a learning rate of recognising actions based on e.g. specific stimuli properties. For this, we compared the accuracies per block with a repeated measures ANOVA. The control-calculations were aimed at evaluating our novel stimulus data set. Finally, we analysed the rating scores per action per context with descriptive statistics (mean and standard deviation), to get insights about the typicality of a given action.

2.2 fMRI Experiment

2.2.1 Design

For the fMRI experiment we altered the experimental design of the precedent behavioural study only slightly to tailor it to, and hence comply with, constraints of the scanner environment. We drastically reduced the number of actions to two, and increased

the number of presentations per action, to reach the necessary amount for the decoding. Finally two actions in two contexts (grasping for the kitchen and hammering for the garage context) were shown. The trials were distributed over 4 blocks, which lasted approximately 7.5 min each. Each action was presented 24 times per block (eight times in every condition), which results in 48 trials per block, so 192 valid trials in total. The order of trials within a block was randomised and per block, four confusion trials were added to keep the participants engaged. The congruent context of action A was simultaneously the incongruent context for action B. To have easy access to the button-box with one hand, we reduced the answer options from four to three after each presented action. Every individual action had a fixed congruent context and a pre-assigned incongruent context. Every context-image was drawn out of a pool of five different images of e.g. a bathroom, to avoid learning effects of the individual image properties.

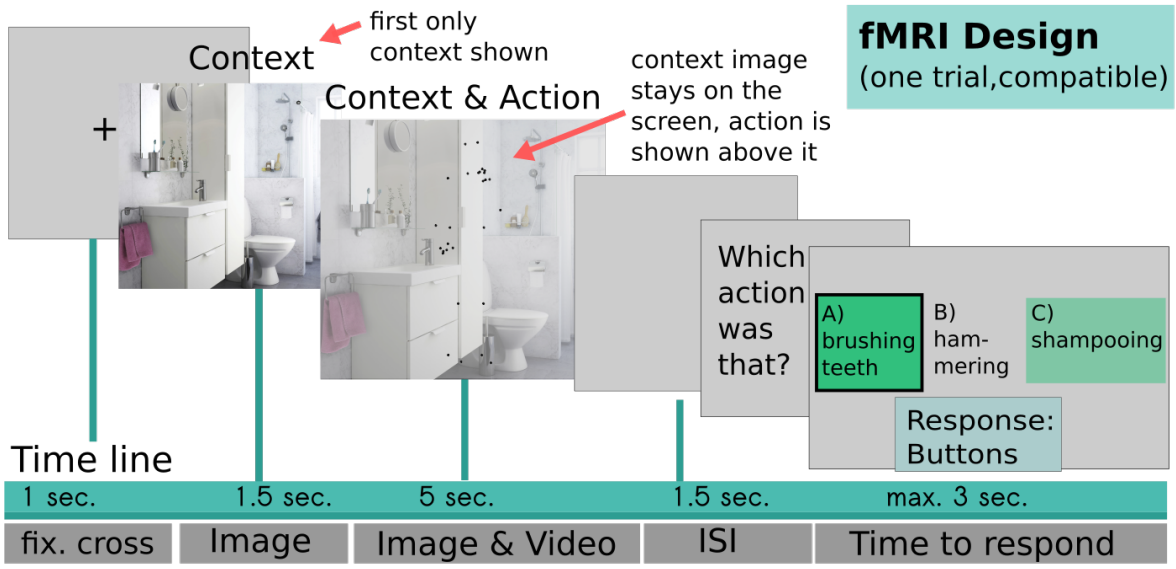


Figure 4: fMRI experiment, altered design for scanner setting

To have an equal number of presented actions in all context conditions in each block, counterbalancing was needed to ensure comparability for the later 'leave-one-block-out' classification analysis. A custom-built script was used to test randomisation of the actions shown in different conditions per block (**uploaded in the repository**). At the forced-choice part, all answer options were shown on horizontally different positions on the screen, so mapped to different fingers. The trial length was 10 sec.

2.2.2 Experimental procedure

We presented the stimuli using PsychoPy version 1.83.03, like in the behavioural experiment. The image-video combinations were projected to a screen at the back of the

scanner, so that the participant was able to see the screen using an adjustable mirror construction attached to the MR head coil. Participants were given a button-box in their dominant hand (right) and were instructed about three valid buttons they could press to signal their chosen answer option. These buttons were pressed with the index, middle and ring finger. The three answer options were presented horizontally on the screen and the screen position was randomised to avoid anticipatory motor action or biases of those aspects. The three answer options corresponded to the three buttons on the participant's button box. The trial started with a fixation cross, after which the context alone was presented for 1.5sec, after which the action sequence was overlayed. The presentation of a grey screen for 1.5 seconds followed. Then the question and answer options were presented, and participants had 3 seconds to indicate their choice by button press.

fMRI stimuli rendering As the training of the classifier for the decoding requires a high number of trials in each category, we optimised the design in a blocked set-up for the available scanner time of 50min. We altered the presentation of the recorded motion-capture stimuli, while keeping the recorded markers at their exact location. This enabled us to analyze action sequences of everyday actions while not having to rely on only one specific action presentation (which would have very narrow low-level features and no variation). With a custom made Matlab script we generated a unique stimulus per action presentation while keeping all points that were recorded in 3D space over the time-course at their exact recorded position. This was done by taking four different subsets of markers, which intersected, two mirroring options and three different sizes of the action sequences, so the actions varied on three dimensions (see 2).

2.2.3 MRI Protocol

For the fMRI study, a 3 tesla Siemens Prisma (Erlangen, Germany) scanner with a 12-channel head coil was used to acquire the data. During the task T2*-weighted echo-planar images (EPI) were acquired, using a multiband multi-echo sequence (TR = 1500ms, TE1 13.40ms; TE2 = 34.42ms; TE3 = 55.44ms, flip angle 75°). Each volume consisted of 84x84x64 isotropic 2.5mm voxels. To relate functional to anatomic space, a structural T1-weighted MPRAGE image was collected for anatomical localization (TR = 2640ms; flip angle: 11 degrees; 0.8mm isotropic voxels)A distortion map (fieldmap) was acquired for spatial corrections (echo time: 4.7ms, 60° flip angle, 2.4x2.4x2mm voxels.

2.2.4 Data Analysis Pipeline (fMRI)

The dependent variable in the fMRI analysis was decoding accuracy. As an initial step we conducted a univariate analysis on a subject- and group-level. After combining the multiple echoes and converting all files to the nifti-fileformat, the individual subject's data was preprocessed with SPM12 (statistical parametric mapping). In comparison to other preprocessing pipelines for univariate brain data analyses, normalization and smoothing were left out to avoid biasing the machine-learning classifier in the later decoding-step. Subsequently, a subject-level General Linear Model (GLM) was set up to relate time-series voxel data to trial data of the action-context presentations in the three conditions incongruent, congruent and neutral over all blocks. On subject-level we used a GLM to relate neural activity to the action-context presentations. We included 3 regressors for context (congruent, incongruent and neuttral), two regressors for actions (grating and hammering) and convolved with HRF response. As a final step on the subject level, we performed univariate statistical tests on neutral vs. context, congruent vs. incongruent and then normalised the results to MNI space to be able to aggregate the data on a group level and perform a univariate group-level analysis.

Multivoxel Pattern Analysis. All decoding steps were performed using 'The Decoding Toolbox' (TDT) Görgen et al., 2018. In order to study the effect of context on action recognition reflected in differential BOLD activity, we trained a machine-learning classifier on the brain activity evoked by one action (e.g. hammering) in congruent trials and tested it on the response of incongruent trials (and neutral trials respectively). Similar comparisons were performed with training the classifier on the activation in neutral trials and testing it on brain activation elicited by incongruent and congruent presentations. To check if we are able to decode context information from the trials in which either garage- or kitchen-images were presented, we used a leave-one-out procedure (blockwise), meaning we trained a classifier on all context only presentations of three blocks and then tested the classifier on the fourth block. This procedure was repeated four times, each time leaving out a different block. The whole-brain searchlight function was used to decode context-information. Adopting an exploratory approach, we used a Whole-brain searchlight function of the decoding toolbox, to check for patterned activity across the whole brain volume.

3 Results

3.1 Behavioural Experiment

First, as described above, we exploratively analysed the accuracies in different conditions per action presented in different contexts. Descriptively, Figure 5 depicts the variability of the context effect depending on the respective action.

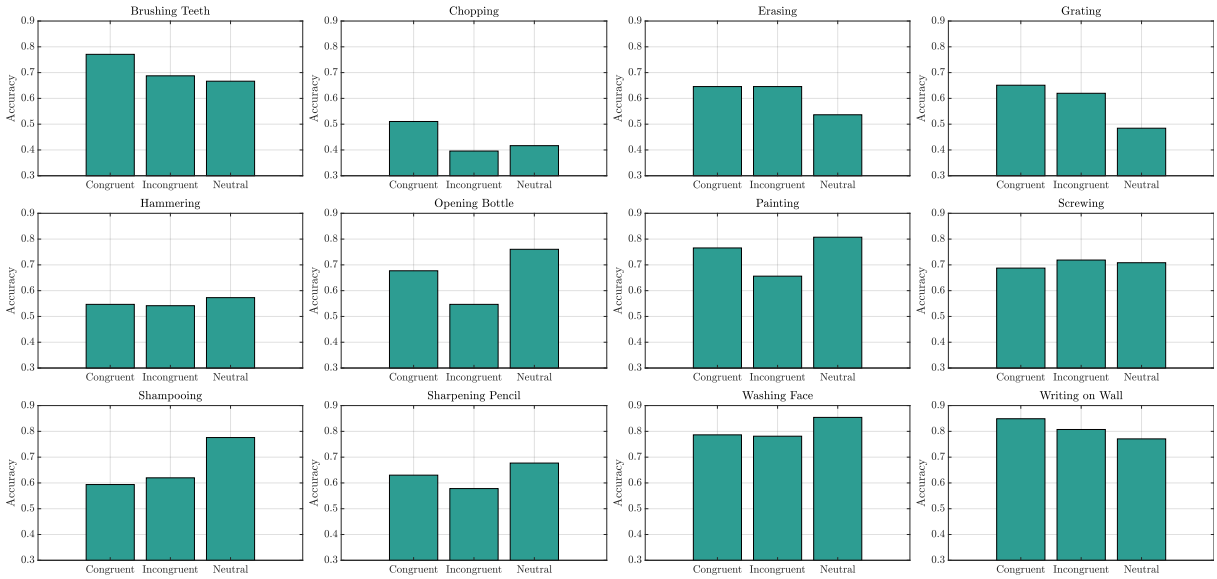


Figure 5: Action accuracies per motion-captured action presented in three conditions

Second, the repeated measures ANOVA with context (congruent, incongruent, neutral) and posture (wide, narrow) as within-subject factors revealed, first, a medium main effect of context-congruency (($F(2,23) = 12.339, p < .001, \eta_p^2 = 0.115$). Hence, the congruency of the context affected the action recognition accuracy.A bonferriori corrected, post-hoc paired-sample t-tests showed significant differences for the incongruent-neutral comparison (t(23)= -4019, p<.001) as well as for the congruent-incongruent (t(23)= 4.538, p<.001) comparison. No significant effect was found for the comparison of congruent-neutral (t(23)= 0.519, p = 1.0). Hence, the incongruent condition has a lower congruency than the neutral and incongruent condition, while the accuracies in the congruent and neutral contexts are not different.

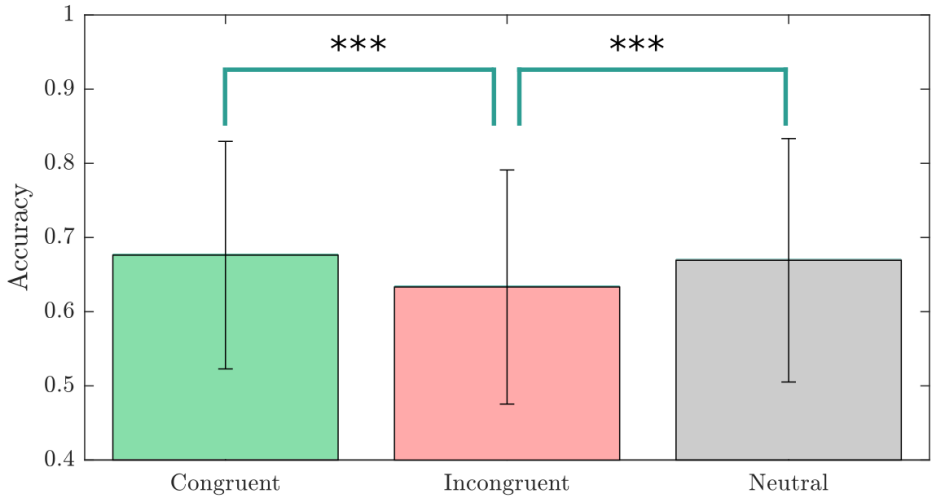


Figure 6: Congruency effect tested with repeated measures ANOVA in Matlab, error bars are standard deviation, significance follows APA standards:* = 0.01 to 0.05 significant, ** = 0.001 to 0.01 very significant, *** = < 0.001 Very significant

Further, the rm ANOVA did not reveal a main effect posture (wide, narrow) ($F(1,23) = 0.009, p = 0.926$. So the adopted posture did not affect accuracies. Finally, contex-congruency by posture interaction did not yield significant results ($F(2,23) = 0.012, p = 0.988$).

The control calculation of accucies over blocks, a rmANOVA with four levels (Block 1 - Block 4), revealed a main effect of learning ($F(3,23) = 10.293, p < 0.001, \eta_p^2 = 0.309$). These accuracies between blocks, which were averaged over participants and all trials as well as conditions in a respective block, therefore significantly differ between the blocks over the course of the experiment. The post-hoc t-tests, which were bonferriori corrected,further showed significant differences between block 1 and block 3 (t(24)= -4.351 p= 0.0001), between block 1 and block 4 (t(24)= -4.008, p= 0.003). Further, the difference between Block 2 and Block 3 is significantly different (t(24)= -3.203, p = 0.024). Hence the accuracy in the first an second block did not change, then from block two to block three, the accuracy significantly increased and stayed constant.

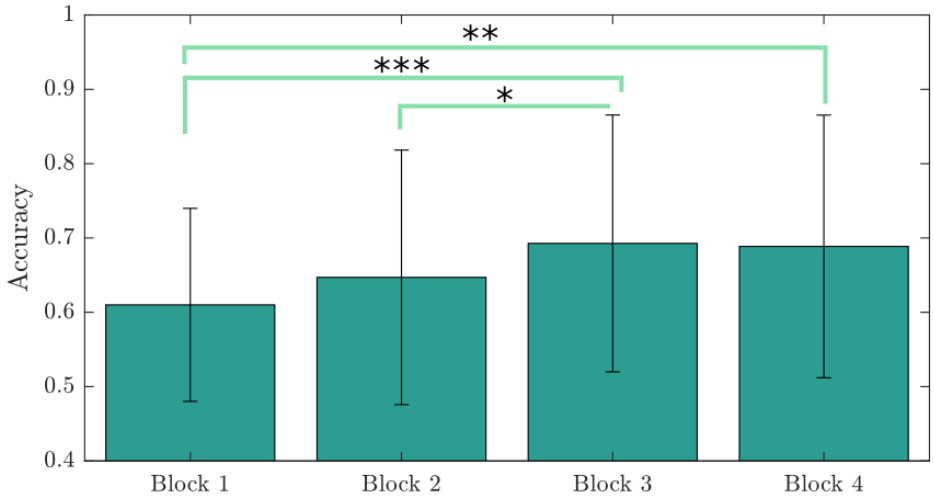


Figure 7: Learning effect tested with a repeated measures ANOVA, Matlab, error bars are standard deviation, significance follows APA standards:* = 0.01 to 0.05 significant, ** = 0.001 to 0.01 very significant, *** = < 0.001 Very significant

The **rating** results are depicted below (8). The heatmap shows the tested actions of the behavioural study in four contexts with the respective means of the rating (N=22). The full numerical data including the standard deviations are to be found in the appendix (1).

Figure 8: Rating results of typicality of action in a given context, N=22

3.2 fMRI Experiment

Due to a multitude of reasons, it was not possible for us to test a whole sample of participants with MRI. This part, however, was an important aspect of my internship and had implications on the experimental design, theoretical considerations as well as the implementation and the programming. Hence, in the following, I discuss expectations

for the fMRI study, but leave out the results, which in their present form are not to be interpreted.

4 Discussion

In the presented thesis work, our goal was to investigate the effects of context information on action recognition. To this end, we designed a new action recognition task and proposed an fMRI decoding paradigm with motion capture actions embedded in a contextual setting. As expected, in the behavioural study we showed that context-congruency has a significant effect on action recognition. The fMRI results are too preliminary to be able to report statistically meaningful results. With the design of a functional MRI decoding task and testing it on first participants, we built on work connecting the neuronal correlates to the established behavioural findings.

4.1 Integration of Results with Previous Research and Theory

Our paradigm diverges from previous studies in three aspects: first, we tested exactly the same motion-captured stimuli in different conditions with a precedent behavioural study. Secondly, we generated unique stimuli yielding unique repetitions of a particular action to use decoding techniques later on without having to rely on one specific video. Finally, our paradigm combines a context-only presentation with a subsequent action-context combination.

Our main finding of congruency effects are in line with the results of Wurm and Schubotz, 2012 who showed that an incongruent context increases the time needed to recognise an action. In Wurm et. al's study, the reaction times of neutral and congruent conditions were not significantly different. In accuracy we showed significant effects of incongruent-congruent and incongruent-neutral settings on action recognition. Notably, congruent context did not elicit a better congruency than the neutral context, only an incongruent context lead to a difference. Hence, the context of the action does not facilitate, but incongruent context information hinders the action recognition.

A few aspects are important to consider regarding the results. Looking at, for instance, the learning rate, the standard deviation shown by the error bars is rather high. This probably is due to the variability between participants. These learning-effects could lead to differences in leave-one-out procedures of the decoding analysis. Having a closer look at the posture-manipulation, it might be important to consider

that the learning effect could be confounding here. Every person had a block in the wide condition at the beginning (either first or second block) and another one in that same condition at the end of the experiment (third or forth, respectively). If there were differences in accuracies, they would mean-out when summed up for the posture manipulation.

4.1.1 Considerations fMRI-Part

We developed our paradigm, hoping to gain insights into how this context involvement in action recognition is represented on a neuronal level, similar to the idea of Amoruso et al. (Amoruso et al., 2016), who approached this with TMS. For the brain activation patterns when processing incongruent contextual information, Wurm et. al (2012) report an increased activation of the ventrolateral prefrontal cortex. This might be an underlying brain region driving the behavioural results we reported. Despite the preliminary state of the neuroimaging data, the presented study shows a high potential for employing decoding techniques to study action recognition in context. I will therefore discuss the expectations we have for a full-sample multivariate pattern analysis, in which we would analyse the time-course decoding accuracies over the time of the trials, similar to (Albers et al., 2018, furhter explained in section 4.2.4).

The ventral premotor cortex (vPM) was shown to be involved in action recognition processes early on (Rizzolatti et al., 2002) as well as in perceiving biological motion (Peelen et al., 2005, Casile et al., 2010). We expect a difference between congruent and incongruent conditions in action decoding at the later stage of the decoding time-window, in which the video was presented (Oosterhof et al., 2012). If decoding of the context was possible in the vPM, either in the beginning when only the context image was presented or over the whole trial, this would be a strong indicator of the importance of context. Similarly, the supplementary motor area as a region-of-interest follows the idea of motor resonance, and we would expect an above chance decoding accuracy of the action there due the brain-regions motion sensitivity in this area. A congruency-effect would be less expected. Looking at scene-information, we expect the Occipital Place Area (OPA) to show context-decoding accuracies above chance, starting as soon as context information is available on screen and getting more reliable with the time the context-image is presented (Kamps et al., 2016). Based on Saxe and Kanwisher, 2003, temporoparietal junction (TPJ) is thought to have an integrative function of different information contributing towards action recognition. The TPJ could therefore be necessary for the last, actual recognition step, and if analysed by its time-course of activation, would show an increased activation later in the recognition process. The demonstration of a congruency effect in this brain area seems unlikely.

4.2 Internal Validity & Resulting Future Directions

4.2.1 Dependent variable(s)

Analysing the behavioural paradigm, we focused on response accuracy instead of response time. We chose this, because the actions were not comparable in difficulty, which would confound the reaction times, as well as the reading time, which differs from action to action as well. This, of course, puts constraints to the amount of information gathered. In future studies, another possible dependent variable on the behavioural level could be the implementation of an eye tracking measure. With this, it would be possible to analyse where participants fixated, and hence, which kinematic or contextual information was sampled. Additionally, the saccades could aid conclusions about the learning of actions, if e.g. participants make less saccades in later blocks, and rather focus on particular statistical properties.

On the level of brain activation, decoding accuracies were the main dependent variable. As another future direction, one would ideally correlate structural MRI, functional MRI, and MEG/EEG-data to obtain both time and spacial specificity of the results. In this new technique, presented recently by Professor Friston's group, dynamic causal modelling (DCM) is used to compare fMRI data, which in a first step is analysed to determine spatially specific neuronal activation changes. In a second step, these results are used as 'location priors' for the analysis of electrophysiology data acquired with the same experimental paradigm (Jafarian et al., 2020 and Wei et al., 2020).

4.2.2 Comparability of Actions

One of the most challenging aspects of the study was working on the comparability of actions. Different features are likely to have contributed to the final difficulty of an action being recognised: One of the them might have been the kinematics, so the positioning of the action movement in 3D space itself, the action's typicality for the given context and how common the movement is.

The actions we chose did vary in their kinematics time-series. Some actions, like 'washing face', for instance, spanned a bigger range in space than others, e.g. 'grating'. For those cases we tried to match similar actions between the four contexts to have a comparable sample. In future analyses, this movement time-series could be statistically analysed and compared. Perhaps, an eigendecomposition of the position of a specific motion-captured point on the dominant hand would be one solution.

Another aspect to consider in future studies is the **typicality** of actions for a given context. Integrating the knowledge of recent studies (for instance Carmo et al., 2020) with our paradigm, the context effect should be higher for closer associated actions in a given context. More explicitly, this would mean that if e.g. brushing teeth for subject A is highly associated with the bathroom context, this would result in a stronger congruency effect than for subject B, who regularly brushes their teeth in the kitchen. In a future step, e.g. the rating results could be taken into account using the individual typicality ratings as a regressor.

4.2.3 Paticulars & Pitfalls of a Decoding Design

Same Analysis Approach To check the neuroimaging design and to make sure pitfalls in counterbalancing and confounding factors are avoided, Görgen et al. designed a tool to inspect the design and analysis pipeline of fMRI studies (Görgen et al., 2018). Particularly important for decoding designs, even more so when the behavioural results revealed differences that could vary systematically with context, the so called 'same analysis approach' should be used to check if all blocks and training vs. test data are comparable.

If we were to have a full functional MRI sample dataset, we would need to test decoding accuracies, that would have been acquired with the described pipeline including the actual decoding step using the decoding toolbox, against chance. For this it has been established recently, that in contrast to t-tests, **permutation tests** are more statistically valid (Haynes, 2015). As explained by Haynes et al. 2015, we would then use a permutation of the label assignment of the samples (congruent, neutral, incongruent, for instance), and with that gather a distribution of chance-level accuracies, against which a given particular accuracy could be tested. The biggest advantage here is, that permutation testing does not necessitate the statistical assumption of normal distribution (Woolrich et al., 2016,Winkler et al., 2016). Further, Haynes et al. (2015) showed false-positive above-chance baseline decoding accuracies for the case of training and test data not actually being independent from each other. This could happen for instance, due to an undetected, small mistake in the design or analysis-pipeline. "In such a case, the whole brain might mistakenly appear to have decodable information about a cognitive variable." (Haynes, 2015). The mentioned Same Analysis Approach (Görgen et al., 2018) is a suitable way to control for such design pitfalls.

4.2.4 Time-Course Decoding of Action Recognition

Building on the results of the presented study, a logical next step would be the research question, whether the time-course of action recognition differs depending on the congruency of the context information. To approach this question, we could study at which stage of recognition the scene information is taken into account by employing 'time-course decoding'(Albers et al., 2018). In this new technique, decoding accuracies are analysed in time bins, allowing for the assessment of temporal dynamics of context-embedded action recognition.

4.3 External Validity & Future Directions

4.3.1 Real-Life Application

In line with Matusz et al., 2018, who emphasised the divergence between complex and interactive real-world environments and the simplified laboratory research that has been conducted over the last decades, finding a meaningful balance for paradigms has been an important challenge. Of course, the degree of which an experimental paradigm is 'embedded' in a given environment varies, and the presented design is only a first step towards more ecologically valid research, combining scanner-constraints with context-information.

In future studies, Virtual Reality scenarios could be employed to study social cognition in a more realistic way, while still being able to manipulate the whole surroundings (Calabròs and Naro, 2020). Going one step further towards ecological validity would be to investigate contextual effects on social cognitive tasks in real-life settings, instead of the laboratory. Such approaches could e.g. make use of portable EEG systems to track brain activity while people are interacting (as, for instance, Toppi et al., 2016).

4.3.2 Significance

Jack & Schyns (2017) refer to the variable context as 'one of the most complex dimensions to test', and precisely because of the wide spectrum of influencing factors, the author stresses the importance of studying those (Jack and Schyns, 2017). The presented study, carries research one step closer to establishing the significance of context-information in cognitive processes, such as action recognition in our case. This could also be seen as a proof of concept to critically evaluate the external validity of

laboratory studies, or even serve as a proxy for future research. Related disciplines could therefore benefit from the discoveries of the proposed studies. Context is under-represented and under-studied, although it is crucial for generalization to ecological, real-life applications.

Expanding beyond the realm of fundamental neuroscientific research, the combination of our results with related studies, discussed and novel, might help to establish a framework to study social cognition in a bottom-up manner. This framework then could also be relevant to investigate clinically relevant aspects of pathologies that are characterised by difficulties in social interaction such as autism spectrum disorder.

List of Abbreviations

(f)MRI	(functional) Magnetic Resonance Imaging
FFA	Frontal Face Area
IFG	Inferior Frontal Gyrus
MEP	motor evoked potentials
OPA	Occipital Place Area
TMS	transcranial magnetic stimulation
TPJ	Temporo parietal Junction
vPM	ventral Premotor Cortex

List of Figures

Declaration of authorship

I hereby declare that I have developed and written the presented master thesis entirely by myself, and have not used sources or means without declaration in the text. Any thoughts from others or literal quotations are clearly marked. This thesis was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.

Anna B.C. Trimborn Nijmegen, 2nd July 2020

Acknowledgements

First, I would like to thank my onsite supervisor Dr Sebo Uithol, who initiated the project and taught me to critically think and question seemingly established concepts. Without him, this internship, one of the richest learning experiences in my academic life, would not have been possible. From recording our stimuli in the baby-lab (between stuffed animals and children's books), to scanner sessions in the middle summer, through the challenges of this project, he was and still is available for advise and support, which I appreciate highly.

I further would like to thank my lab group, the Neuroecology lab headed by Dr Rogier Mars at the Donders Centre for Cognition. I learned so much in the labmeetings about brain analysis, funk-grads and fun-facts about different species! The lab retreat to Oxford was an incredible experience, and I will remember the time in the lab with great fondness and am very thankful for the inspiration and encouragement I experienced.

I in particular also thank Marius Braunsdorf and Dr Suhas Vijayakumar for repeated support during late scanner sessions, but who -more importantly- together with Dr Katherine Bryant and Manon Römkens taught me a lot about academia and life. I appreciated Dr Genevieve Queks input on the low-level feature and Pascal De Waters help with those button-boxes and Python-specifics.

Furthermore, I cannot emphasise enough, how much the scholarships I received contributed to me finishing my degree with this thesis right now. The Friedrich-Ebert-Stiftung, as well as the German Academic Exchange Service enabled me to pursue my studies in a research-focussed masters, continuously funded me through the challenging times of my life, and the effect of this is hard to put into words.

I thank my mother, for equipping me with everything needed to be grounded, but also to grow and discover.

Last, but of crucial importance for sure, I would like to thank the people in my life who -personally and academically- are companions (or: Weggefährten) to me: Thomas Friedrich, Marta Blasco Oliver and Fabienne Windel. You keep me going, and your support is invaluable.

References

Albers, A. M., Meindertsma, T., Toni, I. & de Lange, F. P. (2018). Decoupling of BOLD amplitude and pattern classification of orientation-selective activity in human visual cortex. *NeuroImage*, *180*(February), 31–40. https://doi.org/10.1016/j.neuroimage.2017.09.046

Amoruso, L., Finisguerra, A. & Urgesi, C. (2016). Tracking the time course of top-down contextual effects on motor responses during action comprehension. *Journal of Neuroscience*, *36*(46), 11590–11600. https://doi.org/10.1523/JNEUROSCI.4340-15.2016

Blokpoel, M., Kwisthout, J., van der Weide, T. P., Wareham, T. & van Rooij, I. (2013). A computational-level explanation of the speed of goal inference. *Journal of Mathematical Psychology*, *57*(3-4), 117–133. https://doi.org/10.1016/j.jmp.2013.05.006

Bora, E. & Berk, M. (2016). Theory of mind in major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, *191*, 49–55. https://doi.org/10.1016/j.jad.2015.11.023

Borroni, P., Gorini, A., Riva, G., Bouchard, S. & Cerri, G. (2011). Mirroring avatars: Dissociation of action and intention in human motor resonance. *European Journal of Neuroscience*, *34*(4), 662–669. https://doi.org/10.1111/j.1460-9568.2011.07779.x

Calabròs, R. & Naro, A. (2020). Understanding Social Cognition Using Virtual Reality: Are We still Nibbling around the Edges ? *Brain Sciences*, *10*(17), 10–13.

Carmo, J. C., Martins, F., Pinho, S., Barahona-Correa, B., Ventura, P. & Filipe, C. N. (2020). We see the orange not the lemon: typicality effects in ultra-rapid categorization in adults with and without autism spectrum disorder. *Journal of Neuropsychology*, *14*(1), 154–164. https://doi.org/10.1111/jnp.12176

Casile, A., Dayan, E., Caggiano, V., Hendler, T., Flash, T. & Giese, M. A. (2010). Neuronal encoding of human kinematic invariants during action observation. *Cerebral Cortex*, *20*(7), 1647–1655. https://doi.org/10.1093/cercor/bhp229

Cattaneo, L., Fasanelli, M., Andreatta, O., Bonifati, D. M., Barchiesi, G. & Caruana, F. (2012). Your actions in my cerebellum: Subclinical deficits in action observation in patients with unilateral chronic cerebellar stroke. *Cerebellum*, *11*(1), 264–271. https://doi.org/10.1007/s12311-011-0307-9

Chambon, V., Farrer, C., Pacherie, E., Jacquet, P. O., Leboyer, M. & Zalla, T. (2017). Reduced sensitivity to social priors during action prediction in adults with autism spectrum disorders. *Cognition*, *160*(March), 17–26. https://doi.org/10.1016/j.cognition.2016.12.005

de Lange, F. P., Helmich, R. C. & Toni, I. (2006). Posture influences motor imagery: An fMRI study. *NeuroImage*, *33*(2), 609–617. https://doi.org/10.1016/j.neuroimage.2006.07.017

Ferrari, P. F., Gallese, V., Rizzolatti, G. & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, *17*(8), 1703–1714. https://doi.org/10.1046/j.1460-9568.2003.02601.x

Frith, C. D. & Frith, U. (1999). Interacting minds - A biological basis. *Science*, *286*(5445), 1692–1695. https://doi.org/10.1126/science.286.5445.1692

Gallese, V., Keysers, C. & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*(9), 396–403. https://doi.org/10.1016/j.tics.2004.07.002

Görgen, K., Hebart, M. N., Allefeld, C. & Haynes, J. D. (2018). The same analysis approach: Practical protection against the pitfalls of novel neuroimaging analysis methods. *NeuroImage*, *180*(March), arXiv 1703.06670, 19–30. https://doi.org/10.1016/j.neuroimage.2017.12.083

Haynes, J. D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, *87*(2), 257–270. https://doi.org/10.1016/j.neuron.2015.05.025

Heyes, C. M. & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*(6190). https://doi.org/10.1126/science.1243091

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G. & Mazziotta, J. C. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*(3), 0529–0535. https://doi.org/10.1371/journal.pbio.0030079

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

Jafarian, A., Litvak, V., Cagnan, H., Friston, K. J. & Zeidman, P. (2020). Comparing dynamic causal models of neurovascular coupling with fMRI and EEG/MEG. *NeuroImage*, *216*(March), 116734. https://doi.org/10.1016/j.neuroimage.2020.116734

Jokisch, D., Daum, I. & Troje, N. F. (2006). Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. *Perception*, *35*(7), 911–920. https://doi.org/10.1068/p5540

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N. & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *NeuroImage*, *132*, 417–424. https://doi.org/10.1016/j.neuroimage.2016.02.062

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J. & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience*, *29*(32), 10153–10159. https://doi.org/10.1523/JNEUROSCI.2668-09.2009

Kim, J., Park, S. & Blake, R. (2011). Perception of Biological Motion in Schizophrenia and Healthy Individuals: A Behavioral and Fmri Study. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0019971

Kriegeskorte, N. & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Matusz, P. J., Dikker, S., Huth, A. G. & Perrodin, C. (2018). Are We Ready for Real-world Neuroscience? *Journal of Cognitive Neuroscience*, *31*(3), 327–338. https://doi.org/10.1162/jocn

Ondobaka, S., Newman-Norlund, R. D., De Lange, F. P. & Bekkering, H. (2013). Action recognition depends on observer's level of action control and social personality traits. *PLoS ONE*, *8*(11), 1–9. https://doi.org/10.1371/journal.pone.0081392

Oosterhof, N. N., Tipper, S. P. & Downing, P. E. (2012). Viewpoint (in)dependence of action representations: An MVPA study. *Journal of Cognitive Neuroscience*, *24*(4), 975–989. https://doi.org/10.1162/jocn_a_00195

Peelen, M. V., Downing, P. E., Astafiev, S. V., Stanley, C. M., Shulman, G. L. & Corbetta, M. (2005). Is the extrastriate body area involved in motor actions? *Nature Neuroscience*, *8*(2), 125–126. https://doi.org/10.1038/nn0205-125a

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001

Quattoni, A. & Torralba, A. (2009). Recognizing indoor scenes, In *Computer vision and pattern recognition,2009. cvpr 2009 ieee conference*. https://doi.org/10.1109/cvpr.2009.5206537

Rizzolatti, G. & Fabbri-Destro, M. (2010). Mirror neurons: From discovery to autism. *Experimental Brain Research*, *200*(3-4), 223–237. https://doi.org/10.1007/s00221-009-2002-3

Rizzolatti, G., Rizzolatti, G., Fogassi, L., Fogassi, L., Gallese, V. & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current opinion in Neurobiology*, *12*(2), 149–54. https://doi.org/10.1016/S0959-4388(02)00308-2

Saxe, R. & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Saxe, R. & Baron-Cohen, S. (2006). The neuroscience of theory of mind. *Social neuroscience*, *1*(3-4), 1–9. https://doi.org/10.1080/17470910601117463

Todd, M. T., Nystrom, L. E. & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage*, *77*, 157–165. https://doi.org/10.1016/j.neuroimage.2013.03.039

Toppi, J., Borghini, G., Petti, M., He, E. J., De Giusti, V., He, B., Astolfi, L. & Babiloni, F. (2016). Investigating cooperative behavior in ecological settings: An EEG hy-

perscanning study. *PLoS ONE*, *11*(4), 1–26. https://doi.org/10.1371/journal.pone.0154236

Uithol, S., van Rooij, I., Bekkering, H. & Haselager, P. (2011). Understanding motor resonance. *Social Neuroscience*, *6*(4), 388–397. https://doi.org/10.1080/17470919.2011.559129

Vanrie, J. & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, and Computers*, *36*(4), 625–629. https://doi.org/10.3758/BF03206542

Watson, D. M., Hartley, T. & Andrews, T. J. (2017). Patterns of response to scrambled scenes reveal the importance of visual properties in the organization of scene-selective cortex. *Cortex*, *92*, 162–174. https://doi.org/10.1016/j.cortex.2017.04.011

Wei, H., Jafarian, A., Zeidman, P., Litvak, V., Razi, A., Hu, D. & Friston, K. J. (2020). Bayesian fusion and multimodal DCM for EEG and fMRI. *NeuroImage*, *211*(June 2019), arXiv 1906.07354, 116595. https://doi.org/10.1016/j.neuroimage.2020.116595

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F. & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Winkler, A. M., Webster, M. A., Brooks, J. C., Tracey, I., Smith, S. M. & Nichols, T. E. (2016). Non-parametric combination and related permutation tests for neuroimaging. *Human Brain Mapping*, *37*(4), 1486–1511. https://doi.org/10.1002/hbm.23115

Wokke, M. E., Knot, S. L., Fouad, A. & Richard Ridderinkhof, K. (2016). Conflict in the kitchen: Contextual modulation of responsiveness to affordances. *Consciousness and Cognition*, *40*, 141–146. https://doi.org/10.1016/j.concog.2016.01.007

Woolrich, M. W., Beckmann, C. F., Nichols, T. E. & Smith, S. M. (2016). Statistical Analysis of fMRI Data, In *Fmri techniques and protocols*. https://doi.org/10.1007/978-1-4939-5611-1

Wurm, M. F. & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: Contextual information modulates action recognition. *NeuroImage*, *59*(2), 1551–1559. https://doi.org/10.1016/j.neuroimage.2011.08.038

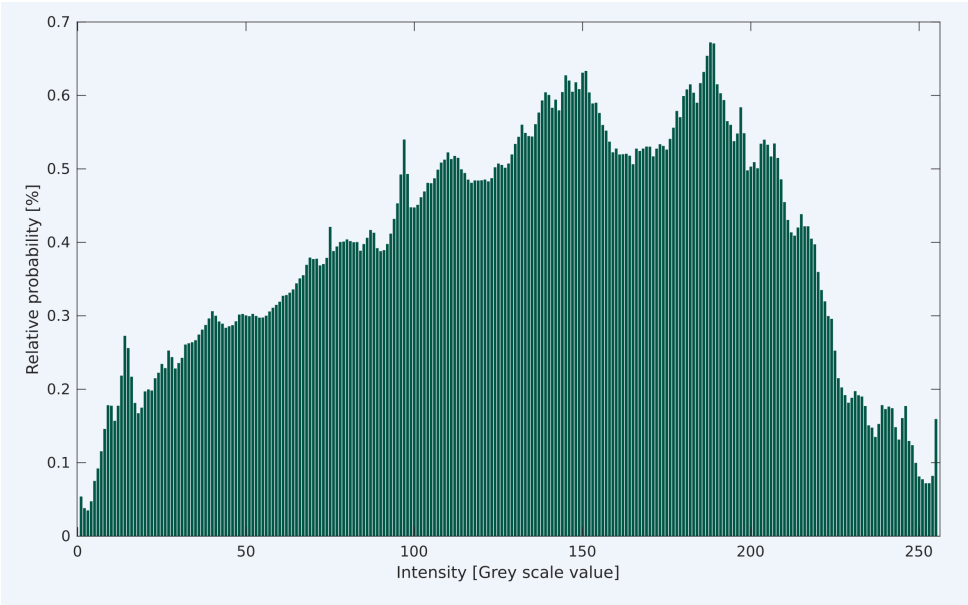
Wurm, M. F. & Schubotz, R. I. (2017). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic Bulletin and Review*, *24*(2), 503–509. https://doi.org/10.3758/s13423-016-1108-4

Zeppi, A. & Blokpoel, M. (2017). Mindshaping the world can make mindreading tractable: Bridging the gap between philosophy and computational complexity analysis. *CogSci*, (July), 1418–1423.

A Appendix

A.1 Unified histogram of greyscale frequency expressions

Figure 9: Unified histogram of greyscale frequency expressions,-the same for every image presented incl. scrambled images



A.2 Mean values and standard deviations of the rating results

Table 1: Mean values and standard deviations of the rating results depicting typicality of actions for a given context, N=22. (see Figure 8 for a graphical representation of this data)

	Bathroom	Office	Kitchen	Garage
Washing Face	6 ± 0	1.1 ± 0.35	2.5 ± 1.4	1 ± 0.21
Shampooing	6 ± 0.21	1 ± 0	1.5 ± 0.67	1 ± 0
Brushing Teeth	6 ± 0.21	1.2 ± 0.43	2.5 ± 1.1	1 ± 0
Sharpening Pencil	1.3 ± 0.46	5.4 ± 0.67	1.6 ± 0.67	2.7 ± 1.3
Erasing	1.9 ± 1.3	3.6 ± 1.9	1.9 ± 0.89	2.2 ± 1.4
Writing Wall	1.1 ± 0.35	5.4 ± 0.66	2.1 ± 1.1	2.3 ± 1
Grating	1.1 ± 0.29	1.3 ± 0.7	5.8 ± 0.66	2 ± 1.4
Opening Bottle	1.7 ± 1.2	1.9 ± 1.2	5.5 ± 0.74	1.8 ± 1.3
Chopping	1.7 ± 1.5	1.1 ± 0.29	4 ± 2.1	4.3 ± 1.8
Painting	2.6 ± 0.9	1.9 ± 0.94	2.7 ± 1.2	5 ± 1.2
Screwing	2.4 ± 0.73	1.8 ± 0.85	3 ± 0.76	5.9 ± 0.29
Hammering	2.4 ± 0.8	1.6 ± 0.9	2.6 ± 0.67	6 ± 0.21