ECAI 2023
K. Gal et al. (Eds.)

doi:10.3233/FAIA230251

Towards a Neuronally Consistent Ontology for Robotic Agents

Florian Ahrens[a;*], Mihai Pomarlan[b], Daniel Beßler[c], Thorsten Fehr[a], Michael Beetz[c] and Manfred Herrmann[a]

[a]University of Bremen, Department of Neuropsychology and Behavioral Neurobiology
[b]University of Bremen, Institute for Linguistics
[c]University of Bremen, Institute for Artificial Intelligence

Abstract. The Collaborative Research Center for Everyday Activity Science & Engineering (CRC EASE) aims to enable robots to perform environmental interaction tasks with close to human capacity. It therefore employs a shared ontology to model the activity of both kinds of agents, empowering robots to learn from human experiences. To properly describe these human experiences, the ontology will strongly benefit from incorporating characteristics of neuronal information processing which are not accessible from a behavioral perspective alone. We, therefore, propose the analysis of human neuroimaging data for evaluation and validation of concepts and events defined in the ontology model underlying most of the CRC projects. In an exploratory analysis, we employed an Independent Component Analysis (ICA) on functional Magnetic Resonance Imaging (fMRI) data from participants who were presented with the same complex video stimuli of activities as robotic and human agents in different environments and contexts. We then correlated the activity patterns of brain networks represented by derived components with timings of annotated event categories as defined by the ontology model. The present results demonstrate a subset of common networks with stable correlations and specificity towards particular event classes and groups, associated with environmental and contextual factors. These neuronal characteristics will open up avenues for adapting the ontology model to be more consistent with human information processing.

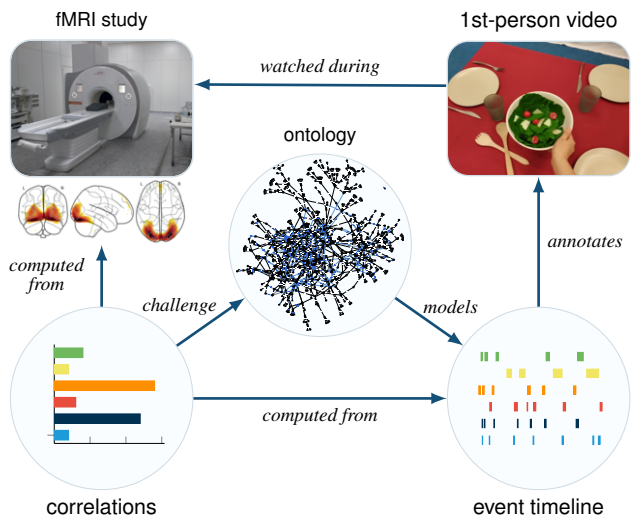


Figure 1. Correlations between neuronal components and event segmentation are used to challenge definitions in an existing ontology.

1 Introduction

The development of autonomous robotic agents by the Collaborative Research Center for *Everyday Activity Science & Engineering* (CRC EASE) is based on the principles of cognition enabled robotic control, employing systems for self-reflected reasoning and planning [8, 7]. Subsystems of the project's cognitive architecture thereby interact with a central knowledge base which is populated not only by the robots' own experiences but also by recorded environmental interactions of humans in the real-word as well as virtual reality contexts [9, 7]. The goal of this effort is to build a knowledge base from experience – whether simulated, observed, or self-performed.

Data in the knowledge base are stored as *Narrative-Enabled Episodic Memories* (NEEMs) which are subdivided into experience and narrative. The NEEM narrative provides a symbolic description of a NEEM in terms of its semantic nature, e.g., the action of picking up a cup, while the NEEM experience contributes the corresponding multimodal sub-symbolic data of the acting agent – human or robot – in the form of, amongst others, audio- and video-recording, motion vectors, captured peripheral physiological parameters or brain activity derived signals. Through this linkage of sub-symbolic and symbolic domains, the robot will be enabled to query task specific experiences, allowing it to adjust the way how an abstract task is executed based on previous experiences of successful executions.

If interoperability between very different systems and/or knowledge transfer between different expert communities is needed, then ontologies are a useful tool towards that goal. An ontology is a collection of axioms in a formal, machine-readable, language which defines terms and makes explicit a conceptualization shared within a community of people. In the case of NEEM narratives, it is the *Socio-physical Model of Activities* (SOMA) ontology [10] which defines the semantic categories which constitute everyday activities and the relationships between entities participating in such activities. As such, SOMA models knowledge about activities both from a robotics perspective, but also aims to model such knowledge as it would be used by humans to organize their own behavior.

* Corresponding Author. Email: fahrens@uni-bremen.de

It is thereby crucial that SOMA, as the basis for the classification of both human and robot behavior and perception, correctly represents and interfaces the experience of both. For the case of robotic agents, this has already been demonstrated [30]. On the human side, a correct representation is arguably ensured by a learned cognitive model that represents the rules for segmentation of everyday behavior into distinctive categories. There is a correlation with the underlying experience stored in dynamically organized networks of brain areas, due to the human ability for self-reflection. However, the process of rule building, itself based on complex brain activity, might introduce different layers of abstraction that obscure aspects of the neuronal activity underlying our experience of everyday activities. It would be beneficial to access the neuronal level to ascertain whether SOMA reflects the underlying processing of these events.

Given the prerequisite as correct that SOMA is considered valid for classifying human experience on a behavioral as well as neuronal level, further analysis directed towards neuronal preferences for certain event classes or levels and a subsequent integration of these neuronal characteristics into the SOMA ontology would be a direct proof of the robot cognitive architecture's being inspired by basic principles of human information processing. We thus decided to investigate the SOMA ontology on a neuronal level via *functional Magnetic Resonance Imaging* (fMRI) to capture brain activity of participants who experienced everyday activities as covered by the SOMA ontology. Since our participants were required to remain in a stationary position inside the MR-scanner, data collection made use of the human brain's ability to simulate environmental interactions.

This procedure refers to the basic assumption that brain activity patterns in the absence of external stimuli through motor imagery [31, 28], follow similar psycho-physical rules as realized during action execution [4, 17] and lead to comparable patterns of brain activation in perceptual [42], as well as motor-related brain networks [18, 44, 48]. If actions are not only imagined but observed, the human equivalent of the mirror neuron system is hypothesized to be recruited as an additional system for mapping observed actions into the subject's own motor representation [43]. Both motor imagery and action observation are thought to arise from a complex neuronal network for learning, maintenance, and refinement of motor actions through a synergy of execution and simulation [22, 29]. This network allows for a close approximation of neuronal correlates of real word-interaction when measuring fMRI participants, especially through use of immersive stimuli with bio-mechanical action representations that have a high degree of ecological validity and correspondence to the observing participants and their motor capabilities [47, 49].

Despite lack of physical engagement in the presented actions during fMRI scanning, it was therefore hypothesized that, given an appropriate level of immersion, brain activity patterns would resemble those present in the real-world experience and therefore allow for a neuronal validation of SOMA ontology through correlation with the spatio-temporal characteristics of its ontological classes as depicted in figure 1. This would contribute to not only SOMA development but ontology development in general by evaluation of an ontology's neuronal validity and potential detection of preferential processing of event categories, enabling new insights for ontology design.

2 Related Work

The work presented here relates to the field of ontology and knowledge engineering as well as to the field of neuroimaging studies during naturalistic viewing. In the following, we will relate our work to the corpus of existing literature in these two fields.

2.1 Knowledge Engineering and Ontology

Information retrieval/organization have motivated ontology creation for medicine and bioinformatics. Examples are the *Neuroscience Information Framework* [26] and the ontologies used for data integration in the *Human Brain Project* [11]. A survey of ontologies for the study of Alzheimer's is provided in [20]. Mercier et al. present steps towards an ontological model, informed by cognitive research, of how a human learns to solve a computational problem [40]; their model is also able to predict learner behavior to some degree.

Ontologies have also been applied to fields such as robotics. As an example, the SOMA ontology [10] defines concepts for autonomous robots to use while performing everyday activities in the home. A broad survey on robotics ontologies is provided by [41].

Some works [37, 2, 6] make a loose distinction between top-down, knowledge-driven and bottom up, data-driven, methods to uncover scientifically-relevant entities. Bottom-up approaches are less affected by expert preconceptions and can recover robust patterns [6], but may confuse dimensionality reduction artifacts with causal mechanisms [37]. Ontologies are developed by a mix of empirical and conceptual issues related to what kinds of entities and questions may be relevant. A summary of this debate can be found in [2] and [37].

Our work here is itself a hybrid top-down/bottom-up approach, in that we start from an ontology developed top-down for robotic actions but compare with human neurological data to ascertain what distinctions between actions humans find relevant.

2.2 Brain Patterns of Naturalistic Viewing

Functional brain imaging has traditionally favoured simple, static stimuli, but the use of complex, dynamic ones may be crucial for analysing the brain in its most natural state (see: [14, 33, 46]). Since the SOMA ontology is meant to describe the dynamic processes carried out by human- and robot agents, neuronal correlates would have to stem from such kinds of stimuli in order to assess its neuronal validity.

Presentation of complex dynamic stimuli during fMRI measurements was shown to be feasible via video with a high degree of inter-participant spatio-temporal correlation of brain activity [21, 12], leading to insights into the neuronal correlates of event and event boundary perception (e.g., [52]). It was further shown that machine learning models could accurately predict the perception of semantic categories from data recorded during video presentation [25].

For initial dimensionality reduction of fMRI recordings, data driven models such as *Independent Component Analysis* (ICA) are used for clustering brain activity into distinct networks through blind source separation [38]. For fMRI data recorded under natural viewing conditions, it was used to subdivide whole-brain activity into spatio-temporal components with characteristic activity time-courses. The association of network activity underlying these components with presented stimuli was analyzed through inter-subject correlation of time-courses, resulting in differentiation of relevant components from non-stimulus related activity and artifacts [5].

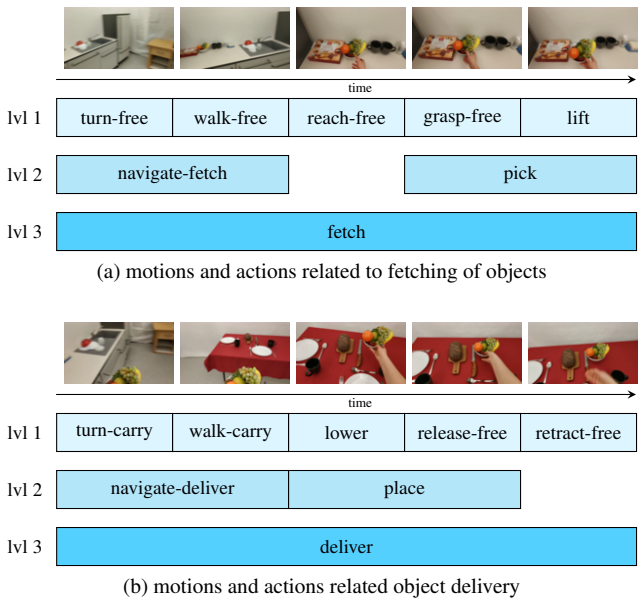
A recording and analysis of neuronal dynamics of everyday activities, which supported the possibility of detecting event dependent allocation of brain activity via *General Linear Model* (GLM) and ICA in the scope of the present research framework was carried out by Ahrens2021 [1]. The study's ICA analysis thereby directly correlated component time-courses with semantically annotated events. Results indicated a set of components common to participants whose activity exhibited such correlations with an additional inter-component

preference depending on the broad classification of the event. These results were however too vague to be used to study human understanding of events.

An updated ICA analysis based on the same data set was thus carried out for the scope of this paper that followed a more stringent ruleset, including multiple comparison correction and a strict focus on stable correlations that were shared amongst participants in order to achieve a more focused set of results.

3 fMRI study

3.1 Experimental Design & Annotation



(a) motions and actions related to fetching of objects

(b) motions and actions related object delivery

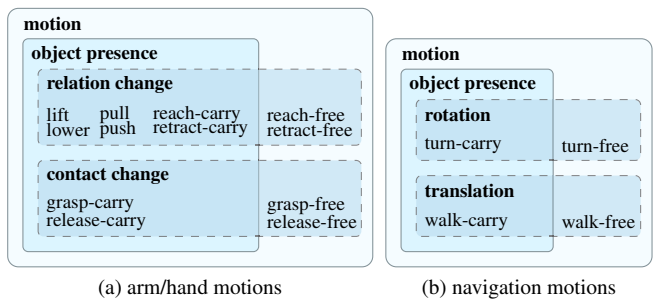
Figure 2. Exemplary visualization of events on annotation levels 1-3.

Stimulus material consisted of ten 1st-person videos of 29 – 105 s duration, recorded via a head mounted camera. As demonstrated in the top rows of figure 2a & 2b, the videos depicted table setting activities in the *Cognitive Systems Lab* (CSL) & the *Institute for Artificial Intelligence* (IAI)) of the EASE CRC at the University of Bremen, integrating the main venues for robot- (IAI) and human- (CSL) activity research into the analysis and representing the project's larger focus on table-setting activities.

The videos covered the fetching of objects from a source area and their subsequent deliverance to a target area for placement in table-setting scenarios. Each scenario was split into a video pair with one part showing the setting of dishes & cutlery and the other part dealing with setting of food & drink items, resulting in four video classes (IAI-dishes, IAI-food, CSL-dishes, CSL-food). Three scenarios were recorded in the CSL and two in the IAI. Both venues differed in factors such as environmental complexity, with the IAI featuring a realistic kitchen environment versus the sparser environment of the CSL, consisting of two tables acting as source and target areas. With the additional split into dish- and food-videos, this allowed for the analysis of contextual effects on measured brain activity and resulting changes in correlation to ontology classes.

Interactions were carried out in a realistic manner, including single- and two-handed movements, while ensuring that all actions remained traceable and recognizable for the viewer. Videos were embedded into an experimental design consisting of two sequences (A & B) of presented table-setting videos and interspersed resting periods. Participants were instructed to watch the table-setting presentations attentively and imagine being the acting protagonist. Both sequences covered ten 1st-person videos, albeit in a switched order with respect to presentation of food & drink and dishes & cutlery parts. Stimuli were presented in a counter-balanced order: half of the study's participants were first presented with sequence A then sequence B, the other half in a reversed order. At the end of a measurement, every participant thus was presented each video twice.

Videos were annotated with EASELAN [39], a modification of the ELAN [36] software by the Max Planck Institute for Psycholinguistics. The process was carried out according to a predefined subset of SOMA events for human activity description developed in collaboration between the human activity and ontology subgroups of EASE CRC. The subset consisted of event categories, nested into annotation levels of increasing length and complexity. Examples of this are depicted in figure 2.



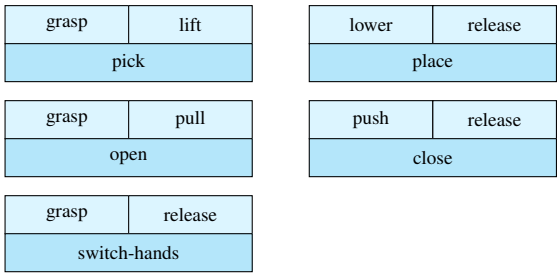
(a) arm/hand motions

(b) navigation motions

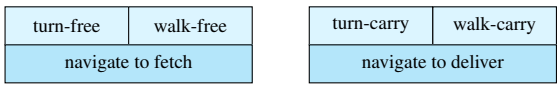
Figure 3. Atomic motion-level annotation scheme.

The lowest annotation level (level 1) consisted of the simplest events in 1st-person videos of everyday activities that are still distinguishable using (combinations of) SOMA motion concepts. These atomic motions are listed in figure 3. They were grouped as either arm/hand motions related to object interaction or motions related to the body's navigation in space. A further distinction was made based on whether an object was held while a motion was performed. This was marked by 'carry' and 'free'. For the motions of grasp and release, it describes whether additional objects were held while performing the respective motion, i.e., grasp-carry indicated that one or more objects were already held while an additional one was grabbed, release-carry indicated that after a release of one object, one or more remained in hand. Other distinctions between motions made by SOMA relate to details of a motion. For example, while 'reach' and 'push' involved similar movements of the hand relative to the body, they involved different force profiles, different contacts, and different states of control over a manipulated object.

The level above atomic motions was the action level (level 2). Actions are performed to reach certain goals and are comprised of a set of motions. For the video annotation we used the actions shown in figure 4. Five action categories were derived from arm/hand motions that covered picking and placing of objects as well as opening and closing of doors and drawers and switching an object between hands. For navigation, we covered two actions. 'Navigate to fetch' represented the navigation motions, usually a rotation followed by a translation, needed to reach the source area, whereas 'navigate to



(a) arm/hand actions



(b) navigation actions

Figure 4. Action-level annotation scheme, including constituting motions.

deliver' covered the respective motions for arriving at the target area.

The level above the action level was the situation level (level 3). For our annotation it consisted of two large-grained time-frames of 'fetch' and 'deliver'. 'Fetch' covered all navigation and object interaction motions that lead to the fetching of item from a source area. During 'deliver', objects were transported to the target area and set down. Examples are shown in the bottom rows of figure 2a & 2b.

Additionally, we included an exploratory group level (level 4) by grouping the atomic motions based on their affiliation to overlapping classes and sub-classes. In most general terms, we grouped motions based on their affiliation to either arm/hand based object interaction motions listed in figure 3a and navigation motions, listed in figure 3b. Within those groups, we subdivided based on 'object presence', describing whether an object was present during motion. For object interaction, we further introduced the categories of 'relation change', describing motions that resulted in change of the spatial relation between hand and body as well as 'contact change', describing motions that either established or broke a contact between hand and object. For navigation motions, we further introduced the categories 'rotation', describing motions in which the body/view is rotated either clock- or anticlockwise and 'translation', describing a forward movement of the body/view in space. Groups were built out of all class affiliations except for navigation motions. Here, no grouping based on object presence was performed since this category was covered by the actions (level 2) of 'navigate to fetch' and 'navigate to deliver'.

Overall, this resulted in a maximum of 16 motion-, 7 action-, 2 situation- and 8 group-categories per video, leading to an overall maximum number of 33 event categories.

3.2 *Participants & Data Acquisition*

Thirty participants (21 identifying as female) with a mean age of 23.3 years (SD = 4.54 years) were recruited from the campus of the University of Bremen. All subjects were right-handed, healthy by own accord, and naïve to the experiment before arriving at the laboratory. Prior to the scan session, participants were given a brief tutorial, including presentation of two short videos similar to those shown during the experiment. All stimuli were shown through a mirror system attached to the head coil. Videos were displayed in a rectangle over a black background. Imaging data were acquired with a Siemens 3 Tesla MAGNETOM Skyra full body scanner. FMRI data were recorded via T2*-weighted multi-band EPI with acceleration factor = 3, TR = 1.1 s, TE = 30 ms, matrix size = 64x64x45 voxels and voxel size = 3x3x3 mm. After completion of the functional recording, a T1-weighted structural scan with matrix size = 255x265x265 voxels and voxel size = 1x1x1 mm was performed.

3.3 *Data Processing & Analysis*

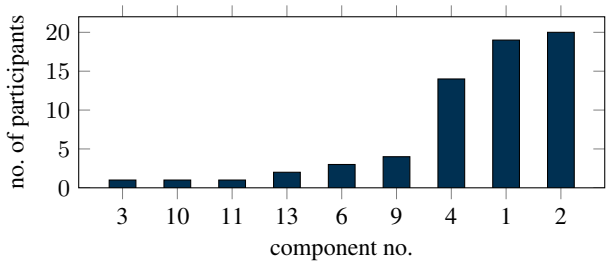
FMRI data were taken from original recordings made for Ahrens2021 [1] which were preprocessed in the *Statistical Parametric Mapping toolbox V12* [15] in *MATLAB V2018b* [27], via subsequent slice-time correction, realignment, coregistration and normalization to standard brain template. Data sets were spatially smoothed with an 8 mm FWHM Gaussian kernel. For each participant, data were separated into blocks temporally corresponding to the presented video- and resting trials via a script taking the *Hemodynamic Response Function* (HRF) into account. For partition of participants' brain activity into spatially independent sources, a spatial ICA was performed on a group level with the *Group ICA of fMRI MATLAB Toolbox V4.0c* [16] over all participants' blocks through an infomax algorithm with an automatic estimation of components numbers. An ICASSO analysis was employed to ascertain the stability of calculated components over repeated calculations [23]. This resulted in classification of brain activity into 15 components common to all participants over all video- and resting trials.

SOMA association of brain activity in component maps was calculated trough Spearman rank correlation between the subject-specific component activity time-course during each video trial and the respective HRF-convolved stimulus timings of the presented sequences of events of a respective category in levels 1-4. Deviating from Ahrens2021 [1], correlation coefficients had to pass a statistical threshold of $p \leq 0.05$, with a Holm-Bonferroni correction [24] for multiple comparisons. The remaining significant correlations further had to proof stable over subsequent presentations of a respective video for each participant. Finally, each stable correlation needed to be found in two or more participants to contribute to the resulting data-set. Components that exhibited such stable and shared correlations were listed and analysed based on their spatial and temporal characteristics.

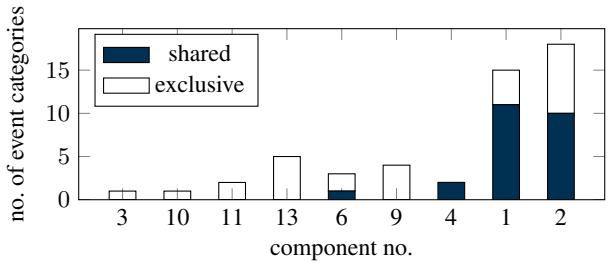
4 Results

4.1 *Grading of SOMA Association*

In nine of fifteen components, at least one out of thirty participants exhibited significant and stable correlations between brain activity within the represented network and at least one annotated event category, as shown in figure 5a. While for components 3, 10, and 11, only one participant was found, two were found in component 13, three in component 6 and four in component 9. The highest numbers of participants were found in component 4 with 14 participants, component 1 with 19 participants and component 2 with 20 participants. The number of event categories with significant and stable correlations for each component is depicted in figure 5b, split into categories exclusive to a particular participant and categories shared between at least two participants. Out of the nine components, only four exhibited shared event categories. In component 6, one such category was found, while two were present in component 4. The highest number of categories were present in component 2 with ten categories and component 1 with eleven categories.



(a) Number of participants with significant correlation of brain activity and annotated events per component.



(b) Number of event categories with significant correlations. Exclusive categories exhibited significant correlation for only a single participant. Shared categories were found in at least two participants A participant could show significant correlation in more than one category for each combination of video and component

Figure 5. SOMA association grading of resulting ICA components.

4.2 *Spatial Brain Activation Maps of Components with shared & stable SOMA Association*

Brain activity patterns associated with component 1 (fig. 6a) were primarily found in the temporal- and occipital lobes, with largest activation in the lateral occipital cortex and extending into the medioventral occipital cortex. In the temporal lobe, the component encompassed the fusiform gyrus. Smaller clusters were also found in the inferior parietal lobule as well as the cerebellum.

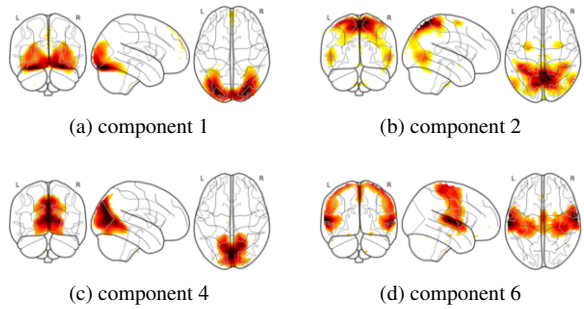
Activity associated with component 2 (fig. 6b) occurred in the parietal lobe with clusters the superior- and inferior parietal lobule, the precuneus and postcentral gyrus. It extended into parts of the lateral occipital cortex, the middle temporal gyrus and inferior temporal gyrus. In the frontal lobe, small clusters were situated in the superior- and middle frontal gyrus, precentral gyrus, and paracentral lobule.

The largest activity cluster for component 4 (fig. 6c) was found spanning areas of the medioventral occipital cortex with smaller clusters in the lateral occipital cortex. It expanded into the superior parietal lobule and the precuneus in the parietal lobe, and the fusiform gyrus and parahippocampal gyrus in the temporal lobe. In the limbic lobe, a small cluster was found in cingulate gyrus.

Activity in component 6 (fig. 6d) was located in the postcentral gyrus of the parietal lobes as well as the superior temporal gyri of the temporal lobe. In the frontal lobe, the component comprised the paracentral lobule and precentral- as well as superior frontal gyri.

4.3 *Intra-Component Correlations*

The number of shared significant stable correlations for all SOMA associated components is depicted in figure 7 via bar graphs for all four video classes. Figure 7a depicts shared stable correlations



(a) component 1 (b) component 2

(c) component 4 (d) component 6

Figure 6. Spatial maps of resulting ICA components that exhibited shared and stable correlations with events of the ontology.

for component 1 ranging over all annotation levels. On level 1, the ICA component correlated with four motion categories (33 correlations in total). Its highest number is for 'reach-free' (14 correlations, 9 CSL-dishes, 5 IAI-dishes), followed by 'grasp-free' (12 correlations, 5 CSL-dishes, 3 IAI-food, 2 IAI-dishes, 2 CSL-food), 'lift' (4 correlations in IAI-food) and 'reach-carry' (3 correlations in CSL-food). On level 2, it correlated with the action category of 'pick' (8 correlations, 4 CSL-food, 4 IAI-food). On level 3, it correlated with the situation category of 'fetch' (15 correlations, 11 IAI-dishes, 4 CSL-dishes). On level 4, it correlated with five exploratory groups (16 correlations in total). Highest correlation is with 'relation change' (6 correlations in CSL-food), followed by 'all arm movement' (4 correlations in CSL-food) and the categories of 'no object present' (2 correlations in CSL-food), 'contact change' (2 correlations in CSL-food) and 'object present' (2 correlations in IAI-food).

Significant and shared stable correlations for component 2 as depicted in figure 7b also covered four annotation levels. On level 1, the component correlated with four motion categories (13 correlations in total). Its highest number was found for 'release-carry' (6 correlations in CSL-dishes), followed by 'lower' (3 correlations in CSL-dishes) and 'retract-carry' (2 correlations in IAI-dishes) as well as 'grasp-free' (2 correlations in CSL-food). On level 2, it correlated with the action category of place (3 correlations in CSL dishes). On level 3, it correlated with the situation category of deliver (5 correlations in CSL-dishes) On level 4, it correlated with four exploratory groups (56 correlations in total). Its highest number found for 'object present' (26 correlations, 19 CSL-dishes, 5 CSL-food, 2 IAI-dishes), followed by 'relation change' (15 correlations, 12 CSL-dishes, 3 CSL-food), 'all arm movement' (8 correlations in CSL-dishes) and 'contact change' (7 correlations, 4 CSL-food, 3 CSL-dishes).

Further significant and stable correlations were found for component 4 in two exploratory groups on annotation level 4 (fig 7d; 19 correlations in total). Both were found for navigation with the highest number for 'navigate rotation' (16 correlations in CSL-dishes), followed by 'navigate all' (3 correlations in CSL-dishes).

In addition, correlations for component 6 as depicted in figure 7d were found on annotation level 4 for the exploratory group 'contact change' (2 correlations in IAI-dishes).

5 Discussion

Out of the four resulting SOMA associated components, brain networks represented by components 1 and 2 had shared and significant stable correlations with ontology events for most participants

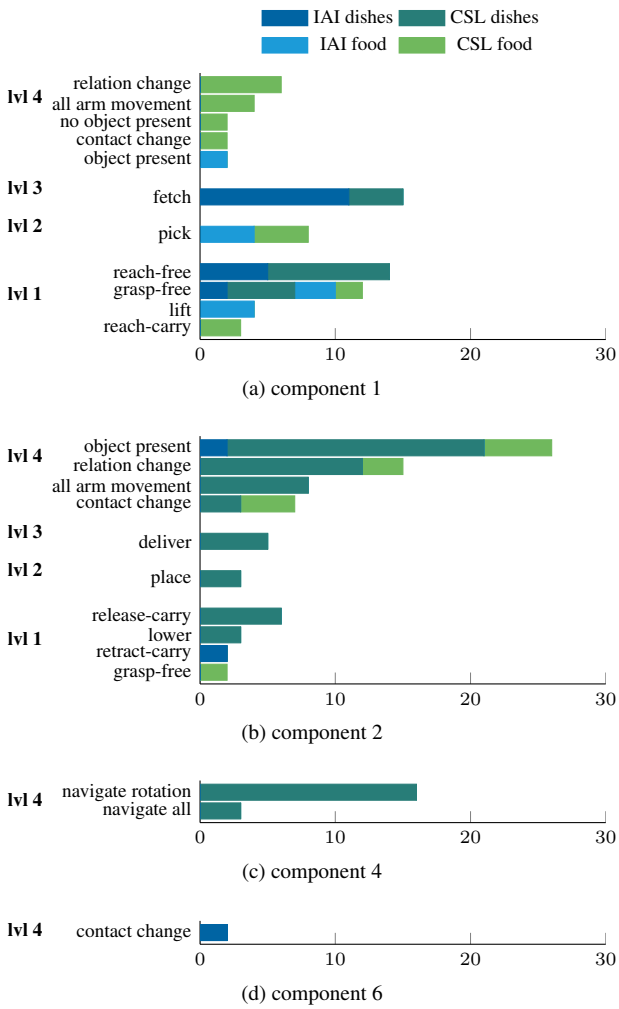


Figure 7. Number of shared and stable correlations for neuronal networks represented by resulting ICA components.

and event classes. For both, associated event classes on annotation levels 1, 2, and 4 were solely based on arm/hand-based object interaction (obligatory navigation involvement in level 3). Out of the minor components 4 and 6, component 4 had a focus on the navigation event of body rotation, while the neuronal network represented by component 6 showed only a minor number of correlations with motions facilitating hand-object contact changes of annotation level 4. These data corroborated the findings of an earlier pilot study by Ahrens2021 [1] for most of the depicted brain networks.

On the basis of significant shared stable correlations and environmental and contextual factors, the current results allow for a distinction of two main networks within four domains, with three of these domains rooted in their characteristics of ontological association.

The first domain, task sensitivity, consists of the functional separation between the concepts of fetch and deliver on event levels 1-3. The network of component 1 thereby representing the concept of fetch and ranging from the situation level (lvl 3) down to its most prominent action (lvl 2) and most constituting motions (lvl 1), while the same holds true in case of component 2 for the concept of deliver.

The second domain, event structure bias, consists of the contrasting bias of both networks towards either single events and situations (lvls 1-3) connected to fetching of objects as found in component 1, or a more pronounced activity during a generalized set of event classes (level 4) as found in component 2.

A third domain, context sensitivity, covers the distinction of findings based on contextual factors. For all components, there was a statistical trend towards correlation to events presented in the context of setting dishes and cutlery as opposed to food and drink items, except for level 4 event grouping of component 1. Out of the two main components, brain activity related to component 2 showed a preference for events recorded in the simpler, less everyday-like environment of the CSL. This was at least in part influenced by the 3:2 ratio of videos recorded in this lab, but further analysis into both factors could prove for a fruitful discussion on this topic.

Underlying all other, a fourth domain of functional characteristics separated networks based on brain activity patterns. Brain activity related to component 1 could be primarily classified into visual perception and object recognition [51, 34], while activation patterns of component 2 covered brain areas that are associated with, e.g., (visuo)spatial attention [3], event boundary perception [52] as well as action planning and execution, including areas hypothesized to house mirror neurons [13, 35]. Additional networks were represented by component 4, which covered areas seen involved in, e.g., whole scene and event boundary perception [50, 52] as well as mental navigation and episodic memory retrieval [19, 32] and component 6 which covered a network involved in primary motor- and somatosensory function and inner verbalization [45]. Associated brain networks thereby give potential explanations for correlation characteristics, e.g., prominent focus on the rotation part of navigation events in component 4 due to its network's potential involvement in perception of boundaries between object interaction and navigation.

They also offer avenues for helping understand the characteristics of the other domains. Examples for components 1 and 2 include: Fetching-related event classes on levels 1-3 are potentially processed on a more perceptual grade compared to those related to delivery which are more closely associated with the planning- and executive network (task sensitivity domain). The planning- and execution network additionally trends towards generalized event classes while the perception-based network focuses on more specialized classes (event structure bias domain). The planning- and execution focused network might furthermore differ in activity based on the environment of video presentation, and both are more active for specific object or event characteristics that are related to dishes & cutlery within the context of everyday activities (environmental sensitivity domain).

We now turn to possible interpretations of the results in relation to validating and improving SOMA. One observation is that the kind of events described by SOMA also correlate well to patterns of observed neuronal activity in human subjects, but more fine-grained conclusions can be obtained based on the domains we described above. The first domain, task sensitivity, validates a particular task distinction made in SOMA, i.e., a functional separation of fetch and deliver concepts.

Meanwhile the third domain, context-related biases, suggests some potential additions to SOMA. While further investigations are required to rule out causes unrelated to brain processing, it appears the brain processes delivery tasks involving dishes differently than delivery tasks involving food. SOMA assumes that the roles played by participants in an activity are specifiable depending on the activity itself, e.g., a delivery task defines a 'patient' role played by the object being delivered. While there are different kinds of patient roles, e.g., cut-object which is a patient role defined by cutting tasks, the attribution of a role to an object is not informed by the type of object nor its

potential subsequent roles in future activities. This ontological commitment might need rethinking, assuming new data from neuronal investigations strengthen the case for revision.

Another potentially important observation is the trending of certain event classes towards a dominance of being processed by networks focused either on perception or planning and execution. This distinction becomes apparent when stable correlations during fetching are compared with ones during delivering events. In the present study, fetching tasks have shown significant stable correlations with perceptual networks, while delivery was preferentially correlated with networks that are related to planning and execution. These results suggest that humans exhibit different neuronal treatment for actions that involve imagined entities such as states which are desired by an agent planning its next action. This fundamental distinction is only covered sparsely by the SOMA ontology. The task taxonomy of SOMA distinguishes between physical and mental tasks where the latter only includes actions whose execution does not involve the agent's body. The ontology further classifies physical tasks based on goals of the agent. However, these task types are not grouped according to whether their execution involves mental entities such as an imagined placement of an object. The results suggest that such a grouping could also be useful in the SOMA ontology. However, further studies will be needed that cover additional action types.

Additional studies would also prove fruitful to gain deeper insights into other characteristics of the current results. For example, only few significant correlations were found with primary motor- and somatosensory networks as represented by component 6, which could have stemmed from the study's focus on stable and shared correlations. Analyses concerning inter-subject variation or changes in correlation strength due to repetition- and learning effects through repeated video presentations might offer valuable insights into underlying neuronal processing characteristics when compared with other networks. Furthermore, since component signals were correlated with events defined by our existing ontology, components could be present in the current data-set whose underlying brain networks were stimulus coupled in novel ways not yet covered by existing ontological classes. Analyses focusing, e.g., on effects of resting periods on component signals could help identifying these networks.

Other avenues for extended analysis concern the preferential correlation to events based on contextual factors. Underlying stimuli difference could thereby be based in both the spatial dimension, e.g., overall environmental complexity or object characteristics, and/or the temporal dimension, e.g., differences in the timing and distribution of scenes and events. An analysis of the stimulus material concerning these factors would thus prove insightful. Finally, to achieve SOMA-related add-on value from the second domain, network bias towards specific vs generalized events, additional knowledge about the temporal nature of both networks is crucial. For example, the planning and execution focused network might exhibit a primarily tonic signal with additional spiking activity during specific object deliverance events for all participants. However, it could also trend towards one or the other for certain groups of participants.

6 Conclusion

The aim of the present study was to validate SOMA with neuronal activity patterns derived from human volunteers who view recordings of actors who conduct several table setting scenarios. This is, to our knowledge, a fairly new line of research and as such there is significant work still to be done, including at a methodological level – e.g., does viewing the same video a second time change which networks get activated? If so, how would that bias stable correlations? As such, we have not gathered data to validate the complete SOMA; however, the early results we have are encouraging. The data derived from ICA and subsequent correlational analyses offer promising insights into domains of human neuronal networks and their relation to SOMA underlying our robot cognitive architecture. Significant and stable correlations were found with various action and situational levels of the ontological concepts. These correlations were also shared between subjects. Furthermore, a functional distinction in the SOMA between the concepts of fetching and delivering were also represented in the neuronal data, substantiating the ontology's validity on an organizational level. Additional neuronal network characteristics not yet present in the ontology were found, which are planned to be integrated into the SOMA concepts, thus bringing it closer to human neuronal information processing.

Acknowledgements

The research reported in this paper has been supported by the German Research Foundation DFG, as part of Collaborative Research Center 1320 EASE - *Everyday Activity Science and Engineering*. We would like to thank the group of Tanja Schultz for allowing access to her lab facilities for video recordings as well as for providing support with EASELAN. Further, we are grateful to Angela Kondinska for assisting during data acquisition.

References

[1] Florian Ahrens, *Neuronal Dynamics of Everyday Activities and their Implications for Robot Control – an fMRI Study*, PhD Thesis, University of Bremen, 2021.

[2] Michael L Anderson, *After phrenology: Neural reuse and the interactive brain*, MIT Press, 2014.

[3] Alfredo Ardila, Byron Bernal, and Monica Rosselli, 'Language and Visual Perception Associations: Meta-Analytic Connectivity Modeling of Brodmann Area 37', *Behavioural Neurology*, **2015**, (January 2015). Publisher: Hindawi.

[4] M. Bakker, F. P. de Lange, J. A. Stevens, I. Toni, and B. R. Bloem, 'Motor imagery of gait: a quantitative approach', *Experimental Brain Research*, **179**(3), 497–504, (May 2007).

[5] Andreas Bartels and Semir Zeki, 'The chronoarchitecture of the human brain—natural viewing conditions reveal a time-based anatomy of the brain', *NeuroImage*, **22**(1), 419–433, (May 2004).

[6] Elizabeth Beam, Christopher Potts, Russell A Poldrack, and Amit Etkin, 'A data-driven framework for mapping domains of human neurobiology', *Nature neuroscience*, **24**(12), 1733–1744, (2021).

[7] Michael Beetz, Daniel Beßler, Andrei Haidu, Mihai Pomarlan, Asil Kaan Bozcuoğlu, and Georg Bartels, 'Know Rob 2.0 — A 2nd Generation Knowledge Processing Framework for Cognition-Enabled Robotic Agents', in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 512–519. 2018 IEEE International Conference on Robotics and Automation (ICRA), (May 2018).

[8] Michael Beetz, Dominik Jain, Lorenz Mosenlechner, Moritz Tenorth, Lars Kunze, Nico Blodow, and Dejan Pangercic, 'Cognition-Enabled Autonomous Robot Control for the Realization of Home Chore Task Intelligence', *Proceedings of the IEEE*, **100**(8), 2454–2471, (August 2012).

[9] Michael Beetz, Moritz Tenorth, and Jan Winkler, 'Open-EASE', in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1983–1990. 2015 IEEE International Conference on Robotics and Automation (ICRA), (May 2015).

[10] Daniel Beßler, Robert Porzel, Mihai Pomarlan, Abhijit Vyas, Sebastian Höffner, Michael Beetz, Rainer Malaka, and John Bateman, 'Foundations of the socio-physical model of activities (soma) for autonomous robotic agents', in *Formal Ontology in Information Systems - Proceedings of the 12th International Conference, FOIS 2021, Bozen-Bolzano, Italy, September 13-16, 2021*, eds., Boyan Brodaric and Fabian Neuhaus, Frontiers in Artificial Intelligence and Applications. IOS Press, (2021).

[11] Ingvild E. Bjerke, Martin Øvsthus, Eszter A. Papp, Sharon C. Yates, Ludovico Silvestri, Julien Fiorilli, Cyriel M.A. Pennartz, Francesco S. Pavone, Maja A. Puchades, Trygve B. Leergaard, and et al., 'Data integration through brain atlasing: Human brain project tools and strategies', *European Psychiatry*, **50**, 70–76, (2018).

[12] Lisa Byrge, Dorit Kliemann, Ye He, Hu Cheng, J. Michael Tyszka, Ralph Adolphs, and Daniel P. Kennedy, 'Video-evoked fMRI BOLD responses are highly consistent across different data acquisition sites', *Human Brain Mapping*, (October 2021).

[13] Jody C. Culham, Cristiana Cavina-Pratesi, and Anthony Singhal, 'The role of parietal cortex in visuomotor control: What have we learned from neuroimaging?', *Neuropsychologia*, **44**(13), 2668–2684, (January 2006).

[14] Simon B. Eickhoff, Michael Milham, and Tamara Vanderwal, 'Towards clinical applications of movie fMRI', *NeuroImage*, **217**, (August 2020).

[15] The Welcome Centre for Human Neuroimaging. Statistical parametric mapping toolbox version: 12 (r7487), 2018.

[16] University Center for Translational Research in Neuroimaging and Data Science. Group ica of fmri toolbox (gift) version: 4.0c (2020), 2020.

[17] Victor Frak, Y. Paulignan, and Marc Jeannerod, 'Orientation of the opposition axis in mentally simulated grasping', *Experimental Brain Research*, **136**(1), 120–127, (January 2001).

[18] Emmanuel Gerardin, Angela Sirigu, Stéphane Lehéricy, Jean-Baptiste Poline, Bertrand Gaymard, Claude Marsault, Yves Agid, and Denis Le Bihan, 'Partially Overlapping Neural Networks for Real and Imagined Hand Movements', *Cerebral Cortex*, **10**(11), 1093–1104, (November 2000).

[19] Olivier Ghaem, Emmanuel Mellet, Fabrice Crivello, Nathalie Tzourio, Bernard Mazoyer, Alain Berthoz, and Michel Denis, 'Mental navigation along memorized routes activates the hippocampus, precuneus, and insula', *NeuroReport*, **8**(3), 739–744, (February 1997).

[20] Alba Gomez-Valadés, Rafael Martínez-Tomás, and Mariano Rincón-Zamorano, 'Ontologies for early detection of the alzheimer disease and other neurodegenerative diseases', in *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pp. 42–50. Springer, (2019).

[21] Uri Hasson, Yuval Nir, Ifat Levy, Galit Fuhrmann, and Rafael Malach, 'Intersubject Synchronization of Cortical Activity During Natural Vision', *Science*, **303**(5664), 1634–1640, (March 2004).

[22] Germund Hesslow, 'Conscious thought as simulation of behaviour and perception', *Trends in Cognitive Sciences*, **6**(6), 242–247, (June 2002).

[23] Johan Himberg, Aapo Hyvärinen, and Fabrizio Esposito, 'Validating the independent components of neuroimaging time series via clustering and visualization', *NeuroImage*, **22**(3), 1214–1222, (July 2004).

[24] Sture Holm, 'A Simple Sequentially Rejective Multiple Test Procedure', *Scandinavian Journal of Statistics*, **6**(2), 65–70, (1979). Publisher: [Board of the Foundation of the Scandinavian Journal of Statistics, Wiley].

[25] Alexander G. Huth, Tyler Lee, Shinji Nishimoto, Natalia Y. Bilenko, An T. Vu, and Jack L. Gallant, 'Decoding the Semantic Content of Natural Movies from Human Brain Activity', *Frontiers in Systems Neuroscience*, **10**, (2016).

[26] Fahim Imam, Stephen Larson, Jeffery Grethe, Amarnath Gupta, Anita Bandrowski, and Maryann Martone, 'Development and use of ontologies inside the neuroscience information framework: A practical approach', *Frontiers in Genetics*, **3**, (2012).

[27] The MathWorks Inc. Matlab version: 9.5.0 (r2018b), 2018.

[28] Marc Jeannerod, 'The representing brain: Neural correlates of motor intention and imagery', *Behavioral and Brain Sciences*, **17**(2), 187–202, (June 1994).

[29] Marc Jeannerod, 'Neural Simulation of Action: A Unifying Mechanism for Motor Cognition', *NeuroImage*, **14**(1), S103–S109, (July 2001).

[30] Sebastian Koralewski, Gayane Kazhoyan, and Michael Beetz, 'Self-specialization of general robot plans based on experience', *Robotics and Automation Letters*, (2019).

[31] Stephen Michael Kosslyn, 'Seeing and imagining in the cerebral hemispheres: A computational approach', *Psychological Review*, **94**(2), 148–175, (1987).

[32] B. J. Krause, D. Schmidt, F. M. Mottaghy, J. Taylor, U. Halsband, H. Herzog, L. Tellmann, and H.-W. Müller-Gärtner, 'Episodic retrieval activates the precuneus irrespective of the imagery content of word pair associates: A PET study', *Brain*, **122**(2), 255–263, (February 1999).

[33] David A. Leopold and Soo Hyun Park, 'Studying the visual brain in its natural rhythm', *NeuroImage*, **216**, (August 2020).

[34] Aleksandar Malikovic, Katrin Amunts, Axel Schleicher, Hartmut Mohlberg, Milenko Kujovic, Nicola Palomero-Gallagher, Simon B. Eickhoff, and Karl Zilles, 'Cytoarchitecture of the human lateral occipital cortex: mapping of two extrastriate areas hOc4la and hOc4lp', *Brain Structure & Function*, **221**(4), 1877–1897, (May 2016).

[35] Monica Maranesi, Luca Bonini, and Leonardo Fogassi, 'Cortical processing of object affordances for self and others' action', *Frontiers in Psychology*, **5**, (2014). Publisher: Frontiers.

[36] The Language Archive Max Planck Institute for Psycholinguistics. Elan version: 6.5), 2023.

[37] Joseph B. McCaffrey and Edouard Machery, 'The reification objection to bottom-up cognitive ontology revision', *Behavioral and Brain Sciences*, **39**, (2016).

[38] Martin J McKeown, Lars Kai Hansen, and Terrence J Sejnowsk, 'Independent component analysis of functional mri: what is signal and what is noise?', *Current opinion in neurobiology*, **13**(5), 620–629, (2003).

[39] Moritz Meier, Celeste Mason, Felix Putze, and Tanja Schultz, 'Comparative Analysis of Think-Aloud Methods for Everyday Activities in the Context of Cognitive Robotics', in *Interspeech 2019*, pp. 559–563. ISCA, (September 2019). Backup Publisher: Interspeech 2019.

[40] Chloé Mercier, Lisa Roux, Margarida Romero, Frédéric Alexandre, and Thierry Viéville, 'Formalizing problem solving in computational thinking: an ontology approach', in *2021 IEEE International Conference on Development and Learning (ICDL)*, pp. 1–8. IEEE, (2021).

[41] Alberto Olivares-Alarcos, Daniel Beßler, Alaa Khamis, Paulo Goncalves, Maki Habib, Julita Bermejo, Marcos Barreto, Mohammed Diab, Jan Rosell, João Quintas, Joanna Olszewska, Hirenkumar Nakawala, Edison Pignaton, Amelie Gyrard, Stefano Borgo, Guillem Alenyà, Michael Beetz, and Howard Li, 'A review and comparison of ontology-based approaches to robot autonomy', *The Knowledge Engineering Review*, **34**, (2019).

[42] Joel Pearson, 'The human imagination: the cognitive neuroscience of visual mental imagery', *Nature Reviews. Neuroscience*, **20**(10), 624–634, (October 2019).

[43] Giacomo Rizzolatti, Leonardo Fogassi, and Vittorio Gallese, 'Neurophysiological mechanisms underlying the understanding and imitation of action', *Nature Reviews Neuroscience*, **2**(9), 661–670, (September 2001).

[44] Muriel Roth, Jean Decety, Monica Raybaudi, Raphael Massarelli, Chantal Delon-Martin, Christoph Segebarth, Stéphanie Morand, Angelo Gemignani, Michel Décorps, and Marc Jeannerod, 'Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study', *NeuroReport*, **7**(7), 1280–1284, (May 1996).

[45] Sukhwinder S. Shergill, Michael J. Brammer, Rimmei Fukuda, Ed Bullmore, Edson Amaro, Robin M. Murray, and Philip K. McGuire, 'Modulation of activity in temporal cortex during generation of inner speech', *Human Brain Mapping*, **16**(4), 219–227, (June 2002).

[46] Erez Simony and Catie Chang, 'Analysis of stimulus-induced brain dynamics during naturalistic paradigms', *NeuroImage*, **216**, (August 2020).

[47] Jennifer A. Stevens, Pierre Fonlupt, Maggie Shiffrar, and Jean Decety, 'New aspects of motion perception: selective neural encoding of apparent human movements', *NeuroReport*, **11**(1), 109–115, (January 2000).

[48] Christoph Stippich, Henrik Ochmann, and Klaus Sartor, 'Somatotopic mapping of the human primary sensorimotor cortex during motor imagery and motor execution by functional magnetic resonance imaging', *Neuroscience Letters*, **331**(1), 50–54, (October 2002).

[49] Yen F. Tai, Christoph Scherfler, David J. Brooks, Nobukatsu Sawamoto, and Umberto Castiello, 'The Human Premotor Cortex Is 'Mirror' Only for Biological Actions', *Current Biology*, **14**(2), 117–120, (January 2004).

[50] Nobuyoshi Takahashi and Mitsuru Kawamura, 'Pure topographical disorientation–the anatomical basis of landmark agnosia', *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, **38**(5), 717–725, (December 2002).

[51] Kevin S. Weiner and Karl Zilles, 'The anatomical and functional specialization of the fusiform gyrus', *Neuropsychologia*, **83**, 48–62, (March 2016).

[52] Jeffrey M. Zacks, Todd S. Braver, Margaret A. Sheridan, David I. Donaldson, Abraham Z. Snyder, John M. Ollinger, Randy L. Buckner, and Marcus E. Raichle, 'Human brain activity time-locked to perceptual event boundaries', *Nature Neuroscience*, **4**(6), 651–655, (June 2001).