Running Head: Coordinating vision and action.

# Coordinating vision and action in natural behaviour: differences in spatiotemporal coupling in everyday tasks

Sharon Scrafton[1,2], Matthew J Stainer[1] and Benjamin W Tatler[1]

[1]University of Aberdeen, UK

[2]Griffith University, Australia

Author Note

This research was conducted as part of the PhD thesis of Sharon Scrafton, funded by the University of Dundee.

Correspondence concerning this article should be addressed to Ben Tatler, School of Psychology, University of Aberdeen, Aberdeen, AB24 3FX, Scotland, UK.

Email: b.w.tatler@abdn.ac.uk or b.w.tatler@activevisionlab.org

Abstract

Vision and action are tightly coupled in space and time: for many tasks we must look at the right place at the right time to gather the information that we need to complete our behavioural goals. Vision typically leads action by about 0.5 seconds in many natural tasks. However, the factors that influence this temporal coordination are not well understood, and variations have been found previously between two domestic tasks each with similar constraints: tea-making and sandwich-making. This study offers a systematic exploration of the factors that govern spatiotemporal coordination of vision and action within complex real world activities. We found that the temporal coordination eye movements and action differed between tea-making and sandwich-making. Longer eye hand latencies, more 'look ahead' fixations and more looks to irrelevant objects were found when making tea than when making a sandwich. Contrary to previous suggestions we found that the requirement to move around the environment did not influence the coordination of vision and action. We conclude that the dynamics of visual behaviour during motor acts are sensitive to the task and specific objects and actions required but not to the spatial demands requiring movement around an environment.

*Key words:* Eye-movements, vision, action, spatiotemporal, coordination.

# Coordinating vision and action in natural behaviour: differences in spatiotemporal coupling in everyday tasks

Real world activities can be reduced to a set of component actions that are linked together in order to produce complex behaviours (Miller, Galanter & Pribram, 1960; Forde & Humphreys, 2002). In many everyday tasks, vision and action have to be coordinated to successfully achieve a visuomotor routine required to complete each subgoal of a task. If we think of attention as a broad network that involves planning, motor action, and vision (Land & Tatler, 2009; Tatler & Land, 2016) then understanding the factors that govern spatiotemporal coordination during the completion of active tasks allows us not only to understand how vision supports our interactions with objects, but also to understand everyday attention. The spatial and temporal coordination of vision and action is surprisingly consistent across a range of real world activities. In space, central vision is directed to the target of the manipulation (Ballard et al., 1992; Land & Tatler, 2009). In time, central vision is directed to the target of the manipulation about 0.5 – 1 second before the hand makes contact (Land et al., 1999; Land & Tatler, 2009). If we can understand the allocation of gaze in space and time around our actions, we may gain valuable insights into the organising principles for human behaviour.

## Understanding visuomotor coordination

If we are to understand the interaction between vision and action in real world behaviour then there is a need to better characterise the relationship between these two systems. Everyday life involves active interaction with our environments as we complete our behavioural goals. Our visual system plays a crucial role in these interactions, supplying information about the state of the world that we interact with and actively seeking the information required for our goals (Findlay & Gilchrist, 2003). The intricate link between vision and action raises questions about the appropriateness of studying vision without action or indeed action in isolation if our goal is to understand the role of these components as they naturally co-occur in behaviour (Kingstone et al., 2008; Land & Tatler, 2009). Epelboim et al. (1995, 1997) showed that for two tasks in which the only difference was the involvement of action (looking at versus tapping coloured pegs) oculomotor behaviour was very different in the presence of action than it was in the absence of action. If we wish understand the interaction between vision and action, then it is important to study this interaction in the context of a complex, everyday activity (Kingstone et al., 2008, Tatler & Land 2016).

**Decomposing everyday activities**

Typically our day-to-day tasks are sequential in nature, with several interrelated but distinct actions, often requiring continuous uptake of visual information. This view of task structure, as a hierarchy of goals and sub-goals was formalized by Schwartz and colleagues (Schwartz et al., 1991, 1995) in a standardised *action coding system* (ACS), intended to be used for assessing brain-damaged patients' performance of everyday tasks. In natural tasks, it is often the case that some elements of the task can be self organized whereas some others are sequentially dependent on each other. The overall task goal of *make tea* is made up of several sub tasks (which themselves are made up of many smaller units of action) and it is possible to self organise the order in which most these are completed. For example, one could start the task in any number of ways by filling the kettle or by retrieving the teapot and adding teabags, or by adding milk to the teacup, and throughout the task it is true that many of the subgoals are flexible in terms of the order of completion. However, there are also several task elements that are sequentially dependant on each other having been completed in a specific order, for example it is necessary to fill an empty kettle before it can be boiled, or to remove the teapot lid before it can be filled. This aspect of tasks clearly requires careful coordination and monitoring in time due to the large number of permutations of the order in which these self-organised actions and sequentially dependent actions can be completed. However, it is not this aspect of coordination that we focus on in the present study but rather that which is required for each of the component object manipulations that are required to complete each task.

Land et al. (1999) extended the hierarchical view of task performance by suggesting that the irreducible unit of behaviour was the *object related act* (ORA), which involved the coordination of vision and action to carry out a manipulation of an object. Understanding how gaze is allocated in space and time to support these individual manipulations can offer important insights into not only the allocation of gaze in natural settings but also the manner in which more complex behaviours may be built up. To date, the factors that govern spatiotemporal coordination of vision and action during object manipulations that occur during real world behaviour have not been investigated systematically. The present study takes this novel approach to understanding the link between vision and action in real world activities, characterising the spatiotemporal coordination of gaze and action at the beginning of object manipulations and the extent to which this is influenced by factors related to the task or to the environment.

**Gaze allocation in space and time during actions**

Completing the sub goal actions that make up part of a main goal, for example switching on the kettle (subgoal) to make tea (main goal), typically require about 3 seconds for completion and involve an average of 5.4 fixations (Land et al., 1999). In terms of the spatial allocation of gaze, most fixations are directed at the object that is the target of the current manipulation (Ballard et al., 1992; Land & Tatler, 2009). Despite there often being multiple fixations, most remain on the target object, and it is not clear whether all of these relocations necessarily provide essential new information about the object (Tatler et al., 2011). Exceptions to this rule are fixations that disengage from the current act temporarily and fixate the target of a future act, before returning to the target of the current act, often referred to as "look-ahead" fixations and occur far more frequently than looks back to previously used objects or looks to irrelevant objects suggesting that they are purposeful and intended (Mennie et al., 2007; Pelz & Canosa, 2001). These look-ahead fixations are thought to serve the planning of ongoing behaviour and appear to help the eye arrive sooner at the object when it is needed for the task (Mennie et al., 2007).

With regard to the temporal allocation of gaze, vision is proactively allocated at the start and end of each subgoal action. For example, when making tea, the eyes are directed to the target about 0.5 – 1 seconds before the object is manipulated (Land et al., 1999). This temporal relationship not only emphasises the proactive nature of gaze allocation in everyday activities but also highlights the importance of understanding gaze allocation in time with respect to our current actions.

The timing of gaze allocation with respect to action is surprisingly consistent across a wide range of real world activities. The eyes tend to lead motor output by about 0.5 – 1 second in activities as diverse as driving (Land & Lee, 1994; Land & Tatler, 2001), music sight reading (Furneaux & Land, 1999), walking (Patla & Vickers, 2003), and reading aloud (Buswell, 1920), but this temporal relationship is not necessarily a fixed one. Within any visuomotor task there is considerable variation in eye-hand latencies (see Land et al., 1999 for an example distribution) and the possible sources of variability have not been explored. Furthermore, this temporal lead of the eyes over the hands is something that seems to develop as new visuomotor skills are acquired (Sailer et al., 2005). These observations raise the question about whether the extent of the lead of the eye over the hands is governed by factors related to the task or environment in which the visuomotor behaviours are studied.

**What factors govern the temporal relationship between vision and action?**

Perhaps the clearest indicator to date that eye-hand latency may be governed by factors such as the task or environment comes from the comparison between previous reports of two domestic tasks. When making tea (Land et al., 1999) or making sandwiches (Hayhoe, 2000) the same overarching principles of visuomotor coordination were evident: with few fixations of task-irrelevant objects, and close links between vision and action in space and time. However, while the eyes typically led the hand in both tasks, there was a considerable difference in the eye-hand latencies between the two studies.  During the sandwich-making task, the average eye hand latency was 0.09 seconds compared with the much longer 0.56 seconds found in the tea making study. Land and Hayhoe (2001) speculated that the difference may arise from the setting in which the tasks were completed. Specifically, when making tea, participants were required to move around the environment to complete the task, whereas when making sandwiches, participants were seated throughout, with all objects within reach. Land and Hayhoe (2001) suggested that the need to move around necessarily imposed a slower tempo for completing the tea making task, with more time between object manipulations and therefore greater opportunity to fixate the next object earlier relative to contact by the hand.

**The present study**

In the present study we considered primarily whether the requirement to move around the environment to complete a task influences the temporal relationship between vision and action, the tendency to look at irrelevant objects and looks to objects that will be used later in the task (a possibility suggested by Land & Hayhoe, 2001). By testing the potential effect of this factor across two tasks previously shown to differ in these measures of visuomotor coordination, we were able to either characterise or rule out the possibility that previously reported differences between these tasks arose from the requirement to move around the environment. Any differences between tasks that remain beyond effects of movement would indicate that visuomotor coordination might be influenced by the task itself or the objects and actions that are required. We used a 2 x 2 repeated measures design, in which participants had to make tea and make a peanut butter and jam sandwich, in conditions where all objects were within reach and in conditions where objects were distributed across two regions of the environment such that the participant to walk around the kitchen in order to complete the task. We consider not only the temporal coordination between vision and action at the start of each action, but also whether task goals or the need to move around an environment influences the frequency of and time spent looking at objects irrelevant to the current task, and how often and for how long gaze is allocated to objects that will be used later in the task (looking ahead to objects that will be needed later). In this way we can gain insights into the factors that govern visuomotor coordination in natural tasks.

In taking the above approach, the present study offers a bridge between truly natural observations of behaviour and more controlled laboratory-based study. We introduced the manipulation of how the objects were distributed around the environment and placed some constraints on the tasks themselves (asking participants to make tea with milk and sugar irrespective of their normal preferences). The study therefore offers a way to explore why previously observed differences between instances of natural behaviour may arise (Land and Hayhoe, 2001) without overly sacrificing the ecological validity of the conditions in which the behaviour is studied (Kingstone et al., 2008).

# Method

## Participants

10 (2 male) Undergraduate Psychology students from the University of Dundee participated in the study in return for course credits. All participants had normal or corrected-to-normal vision. Participants signed informed consent, and the study was approved by the University of Dundee Human Research Ethics Committee.

## Materials

A fully functioning kitchen in the University of Dundee, School of Psychology building was utilised for this study. The room fixtures included worktops and kitchen sinks, shelves and electricity points. Objects required for making tea and sandwiches were laid out in the kitchen along with several distractor objects (Figure 1). Distractor objects included several items typically found in the kitchen, for example a dish sponge, and also several items not typically associated with kitchens, for example a hand held fan. Perishable objects for both tasks were laid out in the kitchen and replaced as and when necessary.

FIGURE 1 ABOUT HERE

## Eye Movement Recording

Eye movements were recorded using the lightweight mobile system manufactured by Positive Science LLC, with two cameras to record the right eye and head-centred view of the scene (See Figure 2). The data from these two cameras were recorded onto camcorders housed in a small lumbar pack worn by participants. Gaze direction was estimated offline using the Yarbus gaze fitting software supplied by Positive

Science, using pupil and corneal reflection tracking. A 9-point calibration procedure was conducted at the start and end of each recording session. Calibration of the gaze model was carried out off-line. Any recording sessions for which a good estimate of gaze could not be obtained from the calibration sequence were discarded. In this way, data were collected at 30 Hz with a spatial accuracy of around 1 degree of visual angle.

FIGURE 2 ABOUT HERE

**Design**

We used a within-subjects design to manipulate two independent variables (task and requirement to move). The two task conditions involved making a cup of tea and making a peanut butter and jam sandwich; participants were instructed to make one of each during each session of the study. Across two sessions, we manipulated the requirement to move by placing all objects for a particular task in one area of the kitchen (requiring no movement during the task) or by spreading the objects over two areas of the kitchen that required the participant to move between these areas during the task (see Figure 1). The arrangement of objects was chosen to provide a natural and plausible distribution of items in a kitchen (avoiding unusual placements that would likely influence viewing behaviour, see Võ & Henderson, 2009), and to ensure as much as possible that objects did not occlude each other (which makes assignment of fixation to objects easier when analysing the collected data). For these reasons we did not change the assignment of objects to locations in the kitchen between participants. The two arrangement conditions were counterbalanced across all participants, with half of the participants performing the tasks in the together condition first and the other half the tasks set in the apart condition first. The two tasks (tea and sandwiches) were counterbalanced for each individual participant's four trials however not across participants. To ensure any results were not due to learning effects, analysis was performed whereby the orders of both task and layout were included as fixed effects in the models (see below).

When the objects were all in one area (the "objects together" condition), we emphasized to participants that they should stand in one location throughout the task and not walk around the kitchen, much the same as if they had been seated on a chair. When the objects were spread several feet apart across two areas (the "objects apart" condition), we explained to participants that they would need to walk around the kitchen to complete the task. When making tea (in both the "objects together" and "objects apart" conditions) participants were asked to use the teapot and make tea with sweetener and milk. Instructions on the tasks were kept consistent across both the "objects together" and "objects apart" conditions, the only difference being whether participants were permitted to move around the room or not. No instruction on the order of subtask completion was given, participants were told to complete the tasks as they normally would, there were no penalties for mistakes (and indeed participants did not make any mistakes in these everyday tasks) and no time pressure.

**Analysis**

Gaze-fitted movies from each recording session were analysed manually on a frame-by-frame basis. Each participant had four associated data files which were analysed, two recordings of tea making (one for each of the together and apart conditions) and two recordings of sandwich making (one for each of the together and apart conditions), no practice trials were performed. For all 10 participants analysed here, data were available for all four tasks (no sessions had to be discarded for any reason), and there were no prolonged periods of data loss during any of the recordings. Within each recording, data loss was minimal because the image of the eye and fit of the model to the pupil were visible in each frame of the data. In this way, any momentary tracker loss due to transient changes in lighting or transient mis-fitting of the gaze model could be identified.

Assignment of gaze to an object was via visual inspection of the data by the coder, with gaze recorded as on an object when the gaze cursor fell within the boundaries of the object. Calibration accuracy in monocular mobile eye tracking data deteriorates with distance in depth from the plane of calibration (known as parallax error). The impact of parallax error was minimised in our study in three ways. First, the Positive Science eye tracker places the scene camera close to the participant's eye (Figure 2A). Second, calibration was performed at arm's length, corresponding to the distance that participants typically view objects from in the chosen tasks. Third, the spacing between objects was such that occlusion and close proximity of objects was rare, making assignment of fixations to objects easier even at distances nearer or further than the calibration plane (see Figure 2B-C for example). Where any ambiguity remained, gaze was assigned to the object most consistent with the upcoming manipulation. Such ambiguities were rare but we did not keep a record of how frequently they occurred.

For all eye movement related measures, we recorded gaze events rather than individual fixations: a gaze event is defined as the time from the first entry to an object to the first exit from that same object, irrespective of the number of fixations made within the object. This coding method further minimises any impact of tracker loss on our data because we needed only to identify the first saccade into and away from an object. Saccades into and away from objects are usually large enough to detect easily from the eye image even in the absence of good tracking by the gaze estimation model, and are usually large enough in amplitude to avoid any ambiguities about whether they are within- or between- object saccades (see Land and Tatler, 2009, Figure 5.6). Thus tacker loss had little impact on our analyses.

Looks to background elements of the environment (walls, ceiling, door) were not coded or analysed.  For each manipulated object we coded the time that gaze was first directed to the object, and the time that the hand made contact with the object. Gaze events to task irrelevant objects were also coded, task irrelevant objects consist of all other objects in the room that are not required for the completion of the task in hand, regardless of whether it was congruent with the environment or not (thus, looks to tea making objects while making a sandwich were classed as irrelevant object looks). Instances where an object later to be used was fixated in advance (i.e. a “look ahead” fixation), without a related action were also coded and analysed.

For analysis of the number of object manipulations completed by participants, looks to task irrelevant objects (number and proportion of looks, and total time spent looking at irrelevant objects), and looks that were directed to objects of future use (i.e. look-ahead gaze events; the number and proportion of looks, and total time spent looking at objects to be used later in the task), data were analysed using 2-way repeated measures ANOVAs with task (tea, sandwiches) and object layout (together, apart) as factors. Partial $\eta^2$ is reported as a measure of significant effect sizes for these analyses. While ANOVAs were appropriate for these DVs, which each summarised behaviour over the course of a trial, for our analyses of the durations of each task irrelevant look, the durations of each look ahead and eye-hand latencies we used Linear Mixed-Effect Models (LMMs) to gain better insights into the factors that govern variation in this measure. For the present study, LMMs have the advantage over ANOVA of modelling across the full dataset and of allowing between-subject and between-item variance to be estimated simultaneously. LMMs were run using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in the R statistical analysis environment (R Core Team, 2014). We calculated p-values for effects by selectively removing the effect of interest and comparing it to the full model using the anova() function in R.

## Results

While the total time for making tea was longer than that for making sandwiches (Table 1), this measure is confounded by the time spent waiting for the kettle to boil (even though it was pre-boiled to minimise waiting time). A better test of the equivalence of the two tasks and two layouts of objects is to consider the number of object manipulations performed in order to complete each task. The number of object manipulations carried out by participants did not vary between the two tasks, $F(1,9) = 1.874$, $p = 0.204$ or the two layouts of objects, $F(1, 9) = 3.212$, $p = .107$ (Figure 3). There was no interaction between the task and layout, $F < 1$. Thus, the two tasks used in the present study were comparable in terms of the number of component actions across the four experimental conditions. The number of manipulations executed to complete each task did not vary between tea and sandwich making, or according to the two layouts of objects. Therefore any differences in the remaining dependent variables must reflect subtle differences in visual behaviour and visuomotor coordination, rather than gross differences in task organisation.

FIGURE 4 ABOUT HERE

**Task irrelevant looks**

There are several ways that we can consider task irrelevant looks. If we compare the raw number of task irrelevant looks, there was a significant effect of both layout, $F(1,9) = 8.496$, $p = .0172$, $\eta^2 = 0.48$ and task, $F(1,9) = 9.324$, $p = .014$, $\eta^2 = 0.51$, with no interaction between the two ($F < 1$, Table 1). Overall more looks to task irrelevant objects were made in the apart condition than in the together condition, and during tea making than sandwich making. However, this may be confounded by the fact that there were more looks at objects overall during tea making, and the apart layout condition (Table 1). It is therefore also informative to consider the looks that were directed at task irrelevant objects as a proportion of all looks. When we considered the data in this way, the results were the same – with effects of layout, $F(1,9) = 16.5$, $p = .003$, $\eta^2 = 0.65$, and task, $F(1,9) = 17.26$, $p = .002$, $\eta^2 = 0.66$.

TABLE 1 ABOUT HERE

A potential confound with using the number (or proportion) of looks is that we are collapsing gaze events that have different durations. When we examined the average time that participants spent looking at task irrelevant objects across the entire task, we find that there was a significant effect of task, $F(1,9) = 6.797$, $p = .028$, $\eta^2 = 0.43$, with more time spent looking at irrelevant objects when making tea. However, there was no effect of layout ($F(1,9) = 2.844$, $p = .1$). When we considered the total time spent looking at irrelevant objects as a proportion of the total time looking at all objects, the effect of task remained significant, $F(1,9) = 8.756$, $p = .016$, $\eta^2 = 0.49$. In this analysis there was an effect approaching significance of layout, $F(1,9) = 3.556$, $p = .09$, in the direction of a greater proportion of time spent on irrelevant objects in the apart condition.

Finally, we can examine the average duration of each of the task irrelevant looks. For this measure, we used LMMs to account for the missing data where some participants made no looks at task irrelevant objects on some trials. We found a significant effect of task, $\beta = 178$, $SE = 70.3$, $t = 2.547$, $p = .027$, with longer gazes to irrelevant objects when making tea, but not of layout and no interaction.

**Looking ahead**

We examined look-ahead gaze events (looks to objects that would be used in the future) in the same way that we analysed task irrelevant looks. Look-aheads here were collapsed across fixations, such that several fixations on a future object would be counted as one look, with the duration summed across fixations. A summary of the data is shown in Table 2. Participants made significantly more look-aheads when making tea than sandwiches, $F(1,9) = 10.5$, $p = .01$, $\eta^2 = 0.51$, but there was no effect of layout and no interaction. The proportion of looks that were directed to objects of future use was also significantly higher for tea than sandwich making, $F(1,9) = 22.14$, $p = .001$, $\eta^2 = 0.67$, with no effect of layout or interaction (both $F$-values $< 1$).

With regards to the total time that participants spent looking ahead, there was no significant interaction (*F*<1*)* but a significant effect of task, $F(1,9) = 20.13$, $p = .005$, $\eta^2 = 0.7$, with more time spent looking ahead when making tea, and an effect approaching significance of layout, $F(1,9) = 4.247$, $p = .069$, in the direction of more time spent looking ahead in the apart condition. When time is considered as a proportion, the effect of task remained significant, $F(1,9) = 20.39$, $p = .001$, $\eta^2 = 0.71$. In this analysis layout had a significant effect, $F(1,9) = 6.349$, $p = .03$, $\eta^2 = 0.42$, with a greater proportion of time spent looking ahead when objects were distributed around the kitchen. There was no interaction ($F < 1$).

Using an LMM, we examined the average duration of look-ahead gaze events. We found that there was a significant effect of layout, with longer durations in the together conditions, $\beta = 400.4$, $SE = 147.9$, $t = 2.708$, $p = .021$. There was no difference in the duration of look-ahead events across tasks, and no interaction.

TABLE 2 ABOUT HERE

**Eye-hand latencies at the start of an action**

An LMM was used to examine the effect of task and layout (and their interaction) on eye hand latency, while controlling for participant as a random effect[1]. Eye hand latencies were calculated as the time between fixating an object and the hand making contact with the object. The maximal model did not converge, therefore the

---

[1] We also ran the models controlling for order effects of task and layout, with the results being the same as when these factors were not controlled for – confirming that learning of the task and environment did not influence eye hand latency behavior.

coefficient between the slope and intercept of the random effect was removed to simplify the model. There was a significant effect of task, $\beta = 0.152$, $SE = 0.049$, $t = 3.133$, $p = .005$, with longer eye-hand latencies when making tea than when making sandwiches. There was no effect of layout, $\beta = 0.03$, $SE = 0.05$, $t = 0.603$, $p = .555$ and no interaction between the two factors, $\beta = 0.072$, $SE = 0.097$, $t = 0.747$, $p = .466$ (Figure 4).

FIGURE 4 ABOUT HERE

One potential confound of the above result is that the objects used in each task were different. Figure 5 shows that there was considerable variance in eye hand latencies across the objects used in each of the tasks. It is therefore unclear whether differences between eye hand latencies in tea making and sandwich making were driven by the subtly different demands of these two domestic tasks, or by the fact that the two tasks involve manipulation of objects whose properties themselves might drive the effect rather than ‘task’ *per se*. To examine whether it was indeed overall task goals, or the objects used in the task that underpinned this apparent effect of task found in the previous LMM, we ran a second LMM that included object as a random effect across layout conditions. In doing so we were effectively able to remove the variance in eye hand latencies that was due to differences between objects. In this LMM the effect of task disappeared, $\beta = 0.127$, $SE = 0.122$, $t = 1.099$, $p = .259$. Confirming the previous LMM, there was still no significant effect of layout, $\beta = 0.027$, $SE = 0.055$, $t = 0.494$, $p = .601$, or interaction, $\beta = 0.045$, $SE = 0.108$, $t = 0.413$, $p = .631$[2].

---

[2] While the second LMM no longer showed an effect of task, this LMM was a better fit to the data than the first LMM, $X^2(4) = 25.448$, $p < .001$, suggesting that eye hand

FIGURE 5 ABOUT HERE

Eye hand latencies therefore seem to vary considerably between objects. However, while the above models suggest this, it is not clear from them why this variation between objects might be observed. Two possible sources that may contribute to this variation are the manner in which objects are manipulated and the physical properties of the objects. While we did not set out to explore these aspects of objects, and thus our stimuli were not selected to address these questions, we can use the natural variation in objects and their uses to make some initial, coarse classifications of objects and compare eye-hand latencies between these groupings of objects.

In order to consider the manner in which objects are used, we compared instances where objects were simply moved from one place to another without any functional manipulation (e.g., moving the teapot from one surface to another for later use) compared to instances in which the object was used for the task (e.g., pouring from the teapot). For this exploration of the results we only included objects that were both moved without specific use and also used during the task. When we compared whether the action being conducted on an object was to move the object, or to perform a use of the object for its function (with reason to move entered as a two-level categorical variable), we found no difference in eye hand latencies, $\beta = -0.178$, $SE = 0.122$, $t = -1.461$, $p = .423$, suggesting that this coarse measure of object use did not result in differences in the temporal coordination of vision and action.

---

latencies are better accounted for by a model that incorporates variation between object than by one that does not.

The natural variation between objects meant that it was hard to explore effects of variation along particular dimensions describing physical properties of objects. However, objects could be grouped according to size in order to consider whether eye-hand latencies differ for objects of different sizes. Indeed, size has implications for object manipulation, with larger objects requiring power grips and smaller ones being graspable using a precision grip; size is also a rough indicator of the likely weight of an object and can thus provide information about grip and lift forces required for object manipulation. We divided objects into groups of large and small (which was necessarily relative to the group of objects as a total) using two different groupings (Table 3), with size treated as a categorical variable with two levels in each case. In the first of these two size comparisons, many of the items were lids, so we removed these and reclassified the objects that were not lids.

TABLE 3 ABOUT HERE

There was a significant effect of our first classification of size, with larger objects being associated with larger eye-hand latencies, $\beta$ = -0.328, *SE* = 0.061, *t* = -5.352, *p* <001. However, when we removed the lid objects that made up most of the small objects in the first classification, we found no difference between our large and small objects, $\beta$ = -0.082, *SE* = 0.061, *t* = -1.334, *p* = .182, although the data did trend in the same direction (see Figure 6). The difference between the two models shows that the apparent effect of size in the first model is removed when lids are not considered. This might imply that lids themselves are associated with different eye hand latencies than other objects (specifically, shorter eye hand latencies), and of course lids are used very differently from the other objects. However, the confound between lids and object size make this result hard to interpret. Finally in our consideration of object size, we examined whether ranking the objects used in the second size comparison by their size was related with eye hand latencies. There was a significant increase in eye hand latency with this course ranking measure, $\beta = 0.018$, $SE = 0.008$, $t = 2.185$, $p = .03$, the direction of which can be seen in Figure 6.

FIGURE 6 ABOUT HERE

Some properties of objects are not fixed, but vary even within the course of a task. In order to consider this aspect of object properties in natural tasks, we compared situations in which vessels were full of liquid to situations in which the same vessels were not full of liquid (using a two-level categorical fixed effect in the model). For the tasks used here, only the bottle, glass, milk, mug and teapot were manipulated in both full and non-full states and so our analyses for this issue was restricted to these objects. An LMM showed that these objects were associated with longer eye hand latencies when they were in their full state than when they were not full, $\beta = -0.391$, $SE = 0.095$, $t = -4.122$, $p < .001$ (Figure 7). It should be noted that for the bottle and milk participants first manipulated the object when full and later manipulated it when not full. In contrast, for the glass, mug and teapot, the object was not full of liquid when first manipulated, but full when manipulated again later in the task. Thus effects of whether or not the object was full of liquid were not dependent upon the order in which these states occurred in the task sequence.

One potential confound for the size effect is that objects that were larger tended to be fluid-containing (e.g., the kettle or teapot). We therefore used whether or not objects could be filled as a factor to re-examine the size effects. When an object's ability to hold liquid was controlled for (by including this as a two level categorical fixed effect in the LMM), the effects of size remained the same. Thus, size effects could not be accounted for by whether objects could contain liquid.

FIGURE 7 ABOUT HERE

## Discussion

We explored the link between vision and action in two complex everyday tasks: making tea and making sandwiches. Specifically we considered whether aspects of visuomotor coordination were sensitive to either the overall behavioural task or the need to move around the environment. We found that the tendency to look ahead of the current action (to targets of future actions), the tendency to fixate task-irrelevant objects in the room, and the temporal coordination at the start of an object related act all differed between the two behavioural tasks studied here. When participants were required to move around the environment in order to complete the task, they more frequently looked at irrelevant objects, spent more time looking at objects that would be used later in the task, and these looks ahead to objects needed in the future were of longer duration. However, we found no effect of the need to move around the environment on eye-hand latencies at the start of each manipulation.

The tendency to look ahead to objects that are the target of future actions was generally in line with but slightly higher than previous reports, which have suggested that around 20% of fixations can be classified as look-aheads when building models, making sandwiches or washing one's hands (Mennie et al., 2007; Hayhoe et al., 2003; Pelz & Canosa, 2001). Our higher proportions an in previous studies may be because we calculated these relative to the total number of gaze events – each comprising potentially more than one fixation – during the task, whereas previous studies calculated proportions relative to the total number of fixations. In our study the frequency of look-ahead fixations was higher when making tea than when making sandwiches. Also sensitive to the task was the prevalence of looks to objects that were task irrelevant. Our results of around 4 – 14% of fixations being directed to irrelevant locations across the four experimental conditions are broadly in line with previous reports of 5% (Land et al., 1999) and 16% (Hayhoe et al., 2003). However, the difference between tasks was opposite in our study than these previous reports: more gazes to task-irrelevant objects were found by the present study when making tea than when making sandwiches. The reasons for the variation in look-ahead fixations and looks to irrelevant objects between tasks is unclear, but it does appear that there are clear differences in these aspects of behaviour in these two domestic activities.

Our findings of more frequent looks at irrelevant objects when the participants needed to walk around the environment is also at odds with the previous comparison of tea and sandwich making discussed by Land and Hayhoe (2001). If the previously observed differences were due to the need to move around when making tea, we would have expected fewer task irrelevant looks when objects were distributed around the kitchen in our study. Our result is perhaps more intuitive: moving around the environment necessarily requires greater exploration of the environment and provides opportunities for off-task fixations while the participant moves between manipulated objects. Thus we might expect from this that more looks to irrelevant objects to occur when the participants were required to move around the environment. A similar argument can be made to explain our finding that participants spent more of their time looking at objects for future use in the task and did so with longer gaze durations when the task-relevant objects were distributed around the environment.

A key aspect of this work was to consider the temporal coordination of gaze within object manipulations. We found eye-hand latencies in our two tasks were broadly in line with the vast majority of previous studies of visuomotor coordination in real world settings, which suggest that vision tends to lead action by between 0.5 and 1 second (Land & Tatler, 2009). While Land and Hayhoe (2001) highlighted that eye-hand latency has been found to differ between the two domestic activities, it was unclear why this was the case, and these authors suggested the likely source of this difference was the environment in which the tasks were conducted and specifically whether or not the participant needed to move around the environment in order to complete the task. Here we can qualify this previous speculation by showing that when the environment and its contents are controlled, we still find differences in eye-hand latencies between tea- and sandwich-making, and these differences are not the result of whether or not participants move around the environment during the task. Thus task goals not only change what we look at (Henderson, 2007; Tatler et al., 2011) but also the temporal coordination of vision and action in natural tasks.

Specifically, we found that eye hand latencies were shorter when making sandwiches than when making tea. It should be noted that while our findings are consistent with the direction of eye-hand latencies found between tea making and sandwich making in previous studies, the mean eye-hand latencies in our experiment were different from those in previous studies. We found latencies 0.55 seconds for sandwich making, which is considerably more than the 0.09 seconds reported by Hayhoe (2000). Our mean eye-hand latency was 0.67 seconds for tea making, which is slightly more than the 0.56 seconds reported by Land et al. (1999). It is not clear why these differences arise between our study and these previous studies. However, our eye-hand latencies in all four experimental conditions fall within the typical range of eye-hand latencies found across a broad range of activities in the real world (see Land & Tatler, 2009).

Prior to the present study it was not clear why this difference between tea and sandwiches had been observed. Land and Hayhoe (2001) speculated that these previously reported differences were likely to have arisen from the fact that participants moved around the environment in the tea-making study, but did not move around for the sandwich-making study. That is, an eye hand latency may become longer as a consequence of a longer approach and reach toward an object. When moving around, this necessarily imposed a slower 'tempo' for the task, with longer between each object related act, and therefore greater opportunity to fixate an object longer before the hand made contact with it. Work from Foerster et al (2011) already raised doubts that task tempo might be the key factor determining variations in eye-hand latency. In a speed cup-stacking task, eye hand latencies were on average 0.42 seconds despite the necessarily high tempo and the fact that all objects were within reach of the seated participants (Foerster et al., 2011). In the present study, we have shown that when the environmental setting was controlled and the need to move around was manipulated, this factor did not influence eye-hand latencies. Thus, the lead of the eye over action was not a consequence of opportunities to look at objects sooner (for example due to more time when approaching the object or longer manual reaches to an object) and was a strategy employed by the visual system irrespective of how objects were distributed around the environment.

At this point in our discussion it is worth considering whether we failed to find an effect of moving around the room on eye hand latencies because this factor does not influence the temporal coordination of vision and action at the start of manipulations or because our study was insufficiently powered. We tested ten participants, which is modest by laboratory-based standards in experimental psychology. However, the time required for manual coding of mobile eye tracking data means that sample sizes in studies of real world eye tracking are often necessarily small. While sample sizes are frequently limited in real world studies, each participant often contributes a large amount of data, which is rich with potential measures that can be coded and analysed. For example, in Land et al.'s (1999) seminal paper on tea making, there were only three participants, but each contributed around three minutes of eye movements and actions for analyses, providing sufficient data to make important contributions to the field. In our study, each participant contributed four task recordings –each making tea twice and making a sandwich twice. Thus the amount of data contributed per participant in our study is high. Of course, as a field, small sample sizes will necessarily limit what can be found in real world studies – small effects are unlikely to be identified in these studies – and raise questions about the generalisability of findings. However, some of these issues can be minimised by departing from traditional ANOVA approaches toward analytical approaches like linear mixed models that not only make better use of data by considering all samples rather than condition means, but also better account for variations between individuals and items within the same model. With regard to the findings of the present study, and our lack of effect of moving around the environment on eye hand latencies, we feel that by using LMMs we minimised the risk of missing genuine effects and it was also the case that the effect of layout was far from significant in our models making the chance that this was due to a lack of power unlikely, especially given significant effects of other factors.

If we accept that our data indicate that differences in eye-hand latencies cannot be explained by the requirement to move around the environment, they must arise from factors related to the tasks themselves. This may be somewhat surprising given that the two tasks are in many ways very similar: both are domestic tasks with similar constraints and the same environment. Of course the two tasks involve different objects and thus related actions, for example in tea making a kettle containing boiling water is a feature whereas in sandwich making there are no items which pose a real personal safety risk if handled incorrectly. Therefore it should be considered that differences between these two tasks arise primarily from the objects that are used and the actions that are carried out with these objects. We found that the difference in eye-hand latencies between our two tasks was due to objects used in each of the tasks, rather than the task itself, raising the possibility that temporal coordination of vision and action may depend upon the properties or uses of the objects themselves.

A variety of lines of evidence suggest that visuomotor interactions with objects may vary depending upon the intended actions with that object and expectations about the properties of that object. When approaching obstacles in a virtual environment, when and where participants directed their eyes to the object depended upon whether they intended to collide with or avoid the object (Rothkopf et al., 2007). When intending to collide with an object in this setting, participants directed their eyes to the centre of the objects, whereas when intending to avoid the object they directed their eyes to the margins of the object. Furthermore, gaze was directed to (and away from) the object sooner when the object was to be avoided, than when it was to be collided with (see also Tatler et al., 2011). Thus intentions to act upon the object varied both the spatial allocation of gaze within the object and the temporal relationship between vision and action. In the present study it was sometimes the case that objects were retrieved from one location and placed on the workspace for use further on in the task compared with times where the object was retrieved for immediate manipulation. Whilst it has previously been demonstrated that the kinematics of grasping depend on the end goal of grasping (Valyear, Chapman, Gallivan, Mark & Culham, 2011) here the instances where objects were either moved to be manipulated later in the task, or actually used for a functional purpose at that moment in time did not produce differences in eye hand latencies and as such this measure of the purpose of object manipulation cannot account for the observed differences in eye-hand latencies found in the present study.

Our actions are influenced by the properties of objects. When we use a precision grip to lift an object, grip and lift forces are scaled to the expected weight of an object we are familiar with (see Flanagan, Bowman & Johansson, 2006 for a review). During reaching for and grasping an object, the aperture of the grasping fingers increases with the size of the object to be grasped (Marteniuk, Leavitt, MacKenzie, & Athenes, 1990). This type of processing occurring prior to touch has also been linked to the amount of information that vision has to extract before physical contact with the object. According to Cole (2008) during the initial part of a fixation to an object about to be manipulated, visual cues about the size, density and weight of an object are being processed. Cole (2008) suggests that even with familiar objects, we visually process the size and combine that knowledge with prior knowledge or expectations about the density of the object. When reaching and lifting familiar objects, on-line visual assessments are completed in order to determine the appropriate acceleration of movement and size-related finger tip force rather than rely on memory (Cole, 2008). It therefore seems likely that such visual analysis of objects is carried out during the period between the eye landing on an object and manual contact with the object. Factors that underlie this visual analysis, such as the size of an object, may therefore underpin some of the variation typically observed in eye-hand latencies during natural tasks, during which not only must this visual analysis be performed, but any adjustments to grip aperture or lift force must also be made as a result of this visual analysis.

The present study used a variety of objects of varying size in each task. Our findings provided some initial evidence that eye-hand latencies may vary with the size of objects, with larger objects being associated with longer eye-hand latencies (although the manner in which object size was expressed led to variation in these findings). Not only does this result suggest that physical properties such as the size of an object may be associated with variation in eye-hand latency, but it also raises questions about what processes are occurring during the period between the eye and hand arriving on an object during a natural task. One possibility is that the initial visual analysis of the object that occurs when the eye arrives is used to judge the size of the object and appropriately adjust the reach to that object (e.g., Cole, 2008). Thus, the difference in eye-hand latency with object size might simply reflect differences in time to reach toward objects depending on their size: a longer and slower reach would result in a longer eye-hand latency. However, if this were the case then the opposite effect of object size should have been observed. Previous research shows that smaller objects are associated with longer and slower reaches (e.g., Berthier et al., 1996; Pryde et al., 1998). It appears therefore that our observed effect of object size on eye-hand latency cannot simply be explained by slower reaches. . While the direction of the effect on eye hand latency suggests that motor planning is an unlikely explanation of our observed effect of object size, it remains that this effect could be perceptual or cognitive in nature. We can make two speculative suggestions about what might underlie the observed effect. First, it may be perceptual in nature, because larger objects can be more easily targeted from further away by the eyes, allowing the eyes to arrive sooner at larger object than smaller ones relative to the time when the hand comes into contact with the object. Second, it may be cognitive; lexical access is faster for names of larger objects (Sereno, O'Donnell & Sereno, 2009). When planning what object is next required in the task, participants may be faster to access stored concepts for larger objects, thus allowing them to initiate search for the object sooner than is the case for smaller objects, for which lexical access is slower. Both of these tentative explanations propose that longer eye hand latencies are the result of being able to get the eyes to the object sooner for larger objects (for a similar argument that longer eye hand latencies reflect factors that allow the eye to arrive sooner see Mennie et al., 2007). A future study that carefully manipulates retinal eccentricity and object size would allow these possibilities to be tested and distinguished.

Whilst typically certain properties of objects are constant such as size, the state of the object often changes during the course of a task. In the present study, several of the objects contained different levels of liquid at different points in the task. We analysed whether this factor influenced eye hand latencies and found that comparing objects when they were full and not full did reveal differences in eye hand latencies, with longer eye hand latencies associated with objects that were full of liquid. Why might fuller objects be associated with longer eye and latencies? If we take the same logic as we did for object size that effects might primarily reflect perceptual, cognitive or motoric aspects of planning, we can speculate that perceptual factors are unlikely to have produced the observed effect. Whether an object is full or not is unlikely to change the conspicuity of an object in peripheral vision, this offering no advantage for perceptual identification and subsequent targeting of full objects. Similarly, our speculation that speed of access for concepts when planning the next action (Sereno et al., 2009) might influence eye hand latency seems unlikely in this case. Rather we suggest that the direction of the observed effect is consistent with an effect of object fullness on motor planning. When a vessel is full of liquid, the vessel will be heavier and will pose more of a risk of spillage than when it is not full. An initial visual analysis of an object when the eye arrives at that object might be sufficient to identify the vessel as being full of liquid and thus trigger a more cautious approach to the object or planning of a larger and smoother lift force when picking up the object. Thus, increased eye-hand latencies for full vessels, may reflect the time required to programme more careful motor approaches to manipulating the object, or may simply reflect the consequently slower manual approach to the object when it is full. This suggestion is consistent with previous studies that have shown that when more careful manipulation of an object is required, such as the preparation of a precision grip (Castiello et al., 1992), or grasping a fragile object (Marteniuk et al., 1987), reaches are slower. It therefore appears that the increase in eye hand latencies that we found when objects were full could be accounted for by slower reaches toward these objects.

Our findings for the effects of object properties on eye hand latency are necessarily preliminary due to the exploratory nature of this aspect of our analyses of two natural everyday tasks. These findings warrant more targeted follow ups in the future, perhaps comparing tasks that share common objects so that we might consider more carefully whether overall task goals influence visuomotor coordination. Similarly, different uses for a single object within a task would allow us to consider whether specific uses of objects influence visuomotor coordination within a common overall task goal.

## Conclusion

The temporal coordination between eye movements and action differed between the two tasks studied here, but did not depend on whether or not the participant was required to move around the environment.  Longer eye hand latencies, more 'look-aheads' and more looks to irrelevant objects were found when making tea than when making a sandwich. The differences between these two tasks seem to arise from the different objects used and this suggests that the properties of the different types of objects used in each task influence the coordination of vision and action. The results for object size and whether or not the object was full of liquid seem to suggest that larger and fuller objects are associated with longer eye hand latencies. These findings suggest that factors relating to the visual analysis of an object and subsequent programming and execution of reach and grasp parameters are important aspects of determining not just our motor limb response but also the spatiotemporal coordination of vision and action.

## References

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. Visual Cognition, 17, 1185–1204.

Ballard, D. H., Hayhoe, M. M., Li, E & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks**.** *Philosophical Transactions of the Royal Society of London B, 337,* 331-339.

Berthier, N. E., Clifton, R. K., Gullapalli, V., McCall, D. D., & Robin, D. J. (1996). Visual information and object size in the control of reaching. *Journal of Motor Behavior*, *28*(3), 187-197.

Buswell, G.T. (1935). *How People Look at Pictures: A Study of The Psychology of Perception in Art*, The University of Chicago Press, Chicago.

Buswell, G.T., (1920). An experimental study of the eye-voice span in reading. *Supplementary Educational Monographs No. 17*.

Cameron BD, Enns JT, Franks IM, Chua R (2009). The hand's automatic pilot can update visual information while the eye is in motion. *Experimental Brain Research 195*:445–454.

Castiello, U., Bennett, K. M., & Paulignan, Y. (1992). Does the type of prehension influence the kinematics of reaching? *Behavioural brain research*, *50*(1), 7-15.

Cole, K. J. (2008). Lifting a familiar object: Visual size analysis, not memory for object weight, scales lift force. *Experimental Brain Research, 188*(4), 551–557.

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*(6-7), 945–978.

Epelboim J, Kowler E, Steinman R, Collewijn H, Erkelens C, Pizlo Z (1995a). When push comes to shove: compensation for passive perturbations of the head during natural gaze shifts. *Journal of Vestibular Research* 5:421–442.

Epelboim J, Steinman R, Kowler E, Edwards M, Pizlo Z, Erkelens C, Collewijn H. (1995b). The function of visual search and memory in sequential looking tasks. *Vision Research, 35*:3401–22.

Epelboim, J., Steinman, R. M., Kowler, E., Pizlo, Z., Erkelens, C.& Collewijn, H. (1997). Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision Research, 37,* 2597-2607.

Findlay, J. M. & Gilchrist I. D. (2003). *Active vision: The psychology of looking and seeing.* Oxford University Press, Oxford.

Findlay, J. M. & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661–674.

Flanagan J.R., Bowman M.C., Johansson R.S. (2006). Control strategies in object manipulation tasks. *Current Opinion in Neurobiology, 16*: 650-659.

Foerster, Carbone, Koesling & Schneider. (2011). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision 11*(7):9, 1–16.

Forde, E.M & Humphreys, G.W. (2002). The cognitive neuropsychology of everyday actions. *Neurocase*. *8*, 59-60.

Foulsham, T. & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8* (2), 6, 1-17.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceeding of the Royal Society of London Series B*, *266*, 2435-2440.

Hayhoe, M. (2000). Vision using routines: a functional account of vision. *Visual Cognition, 7* (1/2/3), 43-64.

Hayhoe, M.M., Shrivastava, A., Mruczek, R. and Pelz, J.B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*, 49-63.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. L. (2007). *Visual saliency does not account for eye movements during visual search in real-world scenes.* In R. van Gompel, M. Fischer, W. Murray, & R. W. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Amsterdam: Elsevier.

Itti, L., Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research., 40*:1489-1506.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition,* **17**, 979–1003.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*, 317–340

Land M. F, Lee D. N., (1994) ``Where we look when we steer'' *Nature (London) 369* 742 - 744.

Land M.F., Tatler B.W. (2009). *Looking and Acting: Vision and eye movements in natural behaviour.* Oxford University Press.

Land, M., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*, 3559-3565.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*, 1311-1328.

Marteniuk, R. G., MacKenzie, C. L., Jeannerod, M., Athenes, S., & Dugas, C. (1987). Constraints on human arm movement trajectories. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *41*(3), 365.

Mennie, N., Hayhoe, M., & Sullivan, B., (2007). Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental Brain Research*. *179*(3), 427-442.

Miller, G. A., Galanter, E. & Pribram, K. H. (1960) *Plans and the structure of behavior.* New York, NY: Holt, Rinehart and Winston, Inc.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117*, 382–405.

Patla A.E., Vickers J.N. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research, 148*:133–138.

Pelz J.B., Canosa R., (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research, 41*(25–26):3587–3596.

Pryde, K. M., Roy, E. A., & Campbell, K. (1998). Prehension in children and adults: the effects of object size. *Human Movement Science*, *17*(6), 743-752.

Rothkopf, C.A., Ballard, D.H.,&Hayhoe, M.M. (2007). Task and context determine where you look. *Journal of Vision, 7(14):16,* 1–20,

Sailer, U., Flanagan, J.R., Johansson R.S., 2005. Eye-hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience. 25*, 8833-8842.

Schwartz, M.F., Montgomery, M.W., Fitzpatrick-De Salme, E.J., Ochipa, C., Coslett, H.B., & Mayer, N.H. (1995). Analysis of a disorder of everyday action. *Cognitive Neuropsychology, 12*(8), 863–892.

Schwartz, M.F., Reed, E.S., Montgomery, M.W., Palmer, C., & Mayer, N.H. (1991). The quantitative description of action disorganisation after brain damage: A case study. *Cognitive Neuropsychology, 8*(5), 381–414.

Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Size matters: Bigger is faster. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1115-1122.

Tatler, B. W. & Land, M. F. (2016). Everyday visual attention. In J. Fawcett, E. Risko, and A. Kingstone (Eds.). Handbook of Attention. (pp. 391–422) MIT Press.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45:*643-659.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5):5, 1–*23*.

Torralba, A., Oliva, A., Castelhano, M., & Henderson, J. M. (2006). Contextual Guidance of Attention in Natural scenes: The role of Global features on object search. *Psychological Review, 113*(4): p. 766-786.

Valyear, K. F., Chapman, C. S., Gallivan, J. P., Mark, R. S., & Culham, J. C. (2011). To use or to move: goal-set modulates priming when grasping real tools. *Experimental brain research, 212(1),* 125-142.

Võ, M. L. H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*(3), 24: 1-15.

Wischnewski, M., Belardinelli, A., Schneider, W.X. & Steil, J.J. (2010). Where to Look Next? Combining Static and Dynamic Proto-objects in a TVA-based Model of Visual Attention. *Cognitive Computation, 2 (4),* 326 – 343.

Table 1. Summary of looking behaviour for task irrelevant objects (standard deviation in brackets). Note that the total time looking at all objects excludes looks to background locations in the scene (walls, ceiling, floor, etc.) and the time spent waiting for the kettle to boil.

| | **Apart Sandwich** | **Together Sandwich** | **Apart Tea** | **Together Tea** |
|---|---|---|---|---|
| **Number of task irrelevant looks** | 4.5 (2.9) | 0.4 (0.5) | 12.5 (12.3) | 3.8 (5.4) |
| **Total looks at all objects** | 43.5 (13.2) | 34.1 (10.6) | 63.9 (22.1) | 49.6 (19.8) |
| **Proportion of looks at task irrelevant objects** | 0.10 (0.04) | 0.01 (0.02) | 0.18 (0.12) | 0.06 (0.06) |
| **Total time looking at task irrelevant objects (s)** | 1.27 (0.9) | 0.10 (0.1) | 6.62 (8.5) | 2.42 (4.9) |
| **Total time looking at all objects (s)** | 80.37 (16.8) | 70.44 (21.6) | 91.79 (42.7) | 95.42 (30.7) |
| **Proportion of time looking at task irrelevant objects** | 0.02 (0.01) | 0.002 (0.003) | 0.07 (0.07) | 0.03 (0.05) |
| **Average duration of task irrelevant looks (ms)** | 297 (123) | 241 (102) | 434 (183) | 489 (271) |
| **Total task duration (s)** | 103.3 (21.0) | 92.3 (19.9) | 130.0 (46.0) | 115.8 (30.5) |

Table 2. Summary of look-ahead gaze events. Total looks can be found in Table 1.

| | Apart Sandwich | Together Sandwich | Apart Tea | Together Tea |
|---|---|---|---|---|
| **Number of look-ahead events** | 10.4 (7.9) | 7.6 (6.6) | 19.2 (10.8) | 16.1 (9.2) |
| **Proportion of looks that were look-aheads** | 0.24 (0.13) | 0.21 (0.13) | 0.39 (0.13) | 0.40 (0.13) |
| **Total time looking ahead (S)** | 4.53 (2.72) | 2.46 (2.04) | 14.12 (10.71) | 9.65 (6.04) |
| **Proportion of time looking ahead** | 0.09 (0.04) | 0.04 (0.03) | 0.25 (0.17) | 0.18 (0.09) |
| **Average duration of look-aheads (ms)** | 1511 (327) | 2073 (637) | 1281 (248) | 1586 (391) |

Table 3. Classifications of size for objects.

| | Objects included in model | |
| --- | --- | --- |
| | Large | Small |
| Size comparison 1 | Bottle, loaf of bread, glass, jam jar, kettle, milk, mug, peanut butter jar, plate, tea caddy, teapot | Coke lid, jam jar lid, knife, milk lid, peanut butter jar lid, sweetener, teabags, tea caddy lid, teapot lid, teaspoon |
| Size comparison 2 | Bottle, glass, kettle, plate, teapot, tea caddy, teapot | Jam jar, knife, milk, mug, peanut butter jar, sweetener, teabags, teaspoon |

## Figure Captions

Figure 1. Layout of the kitchen for all conditions

Figure 2. a) Participant wearing Positive Science eye tracker, b) A frame from the gaze-fitted video of a participant making tea in the objects-apart condition, showing a fixation inside the teapot, c) a frame from the gaze-fitted video of a participant making a sandwich in the objects-together condition, showing a fixation on the peanut butter in the participant's right hand.

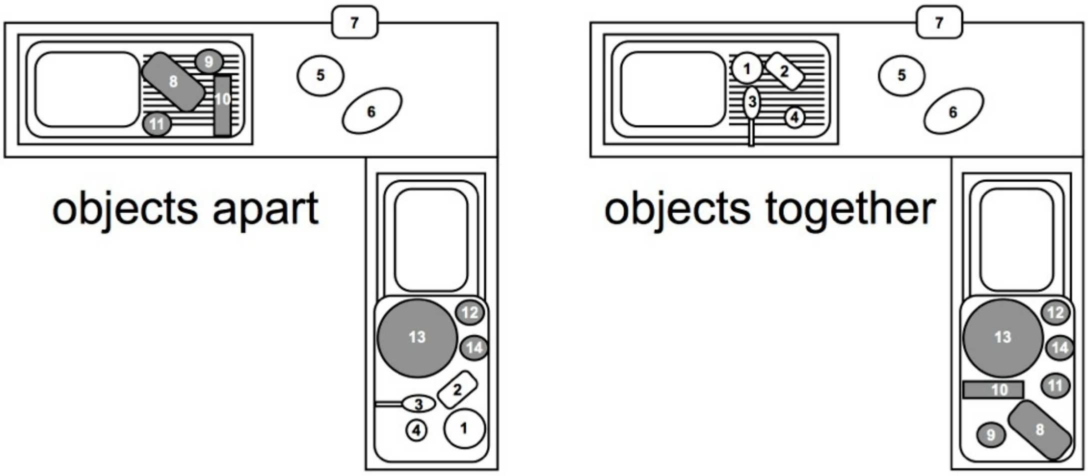
Figure 3. Mean number of object manipulations across tasks in layouts with ±1SEM bars.

Figure 4. Mean and 1±SEM of median eye hand latencies from all participants.

Figure 5. Mean and 1±SEM of median eye hand latencies across all participants separated by object.

Figure 6. Eye hand latencies on objects ordered approximately by size. The objects included in each comparison can be seen in the pairs of bars above, with A for comparison 1 across all objects (with the arbitrary cut-off between Mug and Milk), and B for comparison 2 when lid items were removed (with the cut-off between the Peanut Butter jar and the glass). When all ranked objects were included this involved all of the objects across comparison B (i.e., all objects larger than a Peanut butter jar lid).

Figure 7. Eye hand latencies for objects in their full and not-full state.




| object key | |
|---|---|
| 1. teapot | 8. bread |
| 2. milk | 9. jam |
| 3. teaspoon | 10. knife |
| 4. sweetener | 11. cola |
| 5. mugs | 12. glass |
| 6. kettle | 13. plate |
| 7. tea | 14. peanut butter |

FIGURE 1.

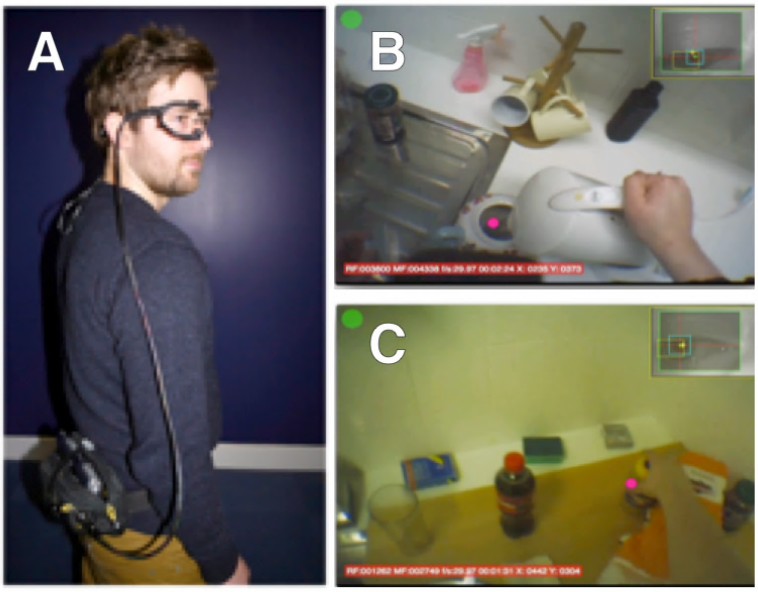



FIGURE 2.

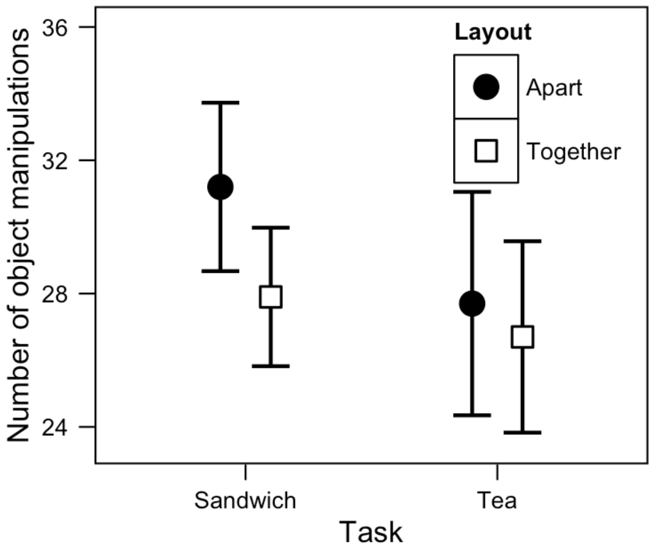

36
32
28
24
Number of object manipulations
Layout
Apart
Together
Sandwich
Tea
Task

FIGURE 3.

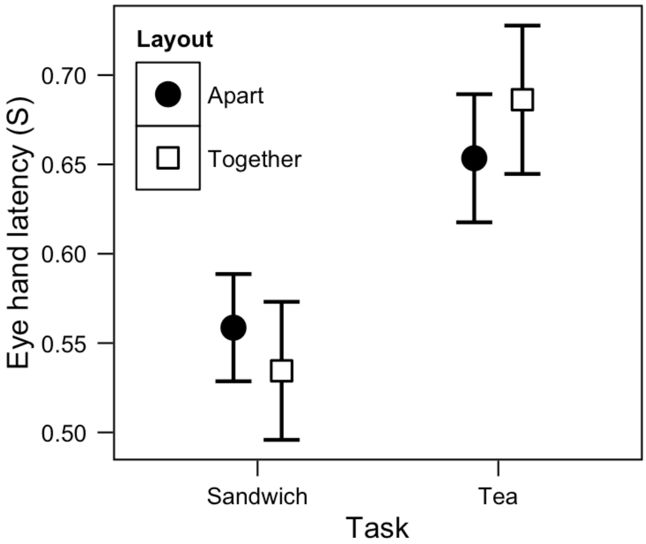

Layout
Apart
Together
Eye hand latency (S)
0.70
0.65
0.60
0.55
0.50
Sandwich
Tea
Task

FIGURE 4.

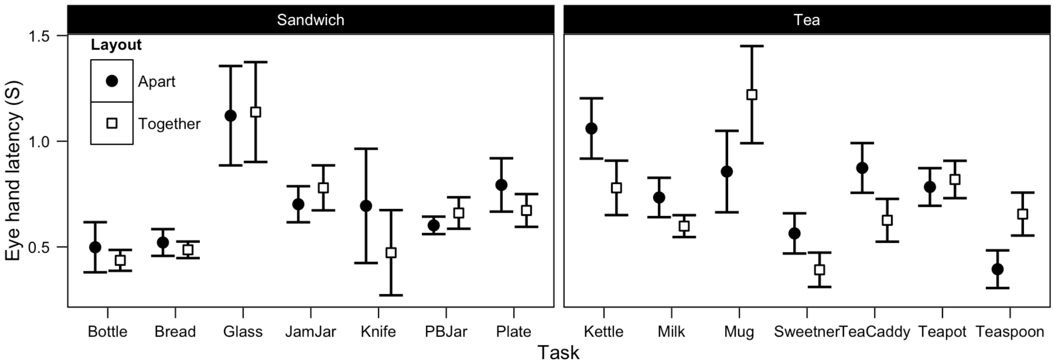

Sandwich
Tea
Layout
Apart
Together
Eye hand latency (S)
1.5
1.0
0.5
Bottle
Bread
Glass
JamJar
Knife
PBJar
Plate
Kettle
Milk
Mug
Sweetner
TeaCaddy
Teapot
Teaspoon
Task

FIGURE 5.

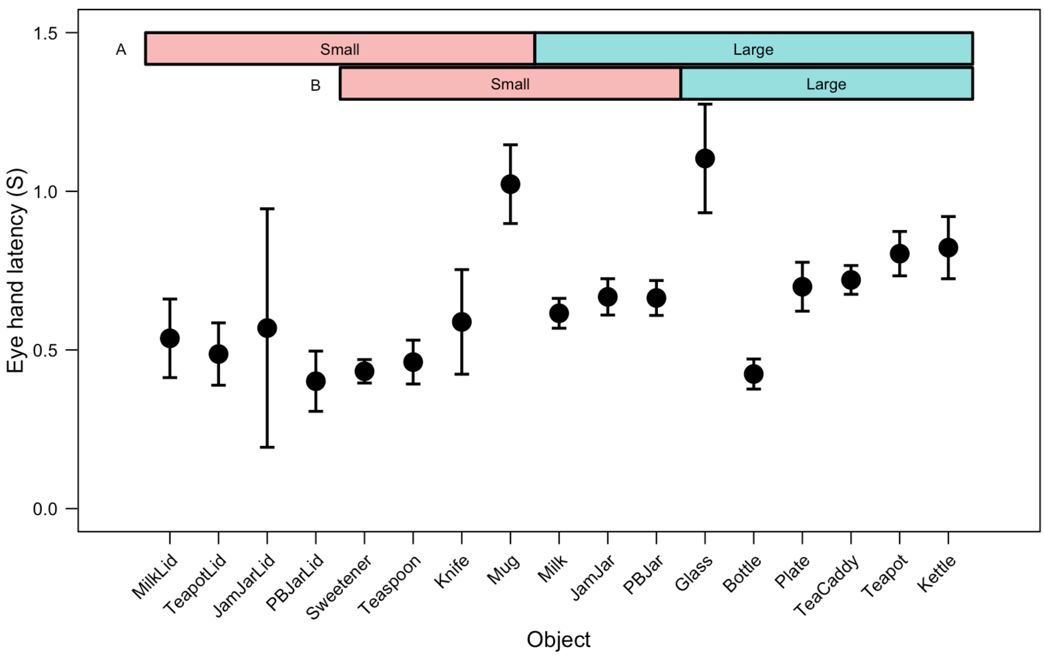

1.5
1.0
0.5
0.0
Eye hand latency (S)
A
Small
Large
B
Small
Large
MilkLid
TeapotLid
JamJarLid
PBJarLid
Sweetener
Teaspoon
Knife
Mug
Milk
JamJar
PBJar
Glass
Bottle
Plate
TeaCaddy
Teapot
Kettle
Object

FIGURE 6.

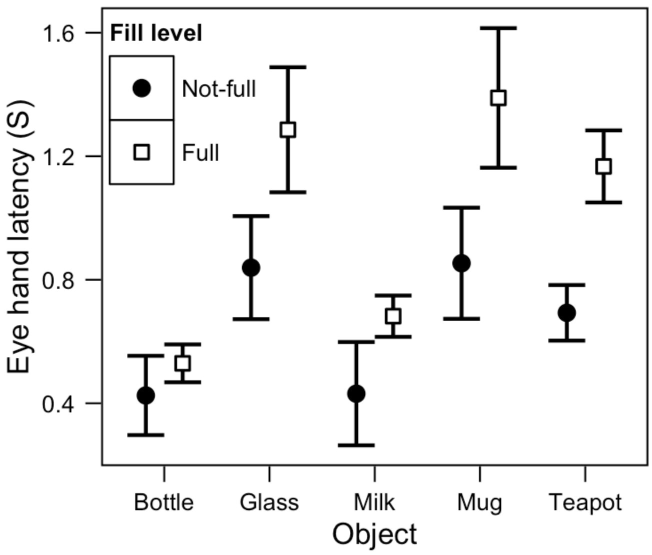

Fill level
Not-full
Full
Eye hand latency (S)
1.6
1.2
0.8
0.4
Bottle
Glass
Milk
Mug
Teapot
Object

FIGURE 7.