The Roles of Vision and Eye Movements in the Control of Activities of Daily Living

Authors: Michael Land, Neil Mennie, Jennifer Rusted

Presenters: Lance Bassett, Nathaniel Chong

Agenda

- Introduce Paper
  - Definitions and Goals
- Experiment
- Findings
  - Time relationships of vision and motor acts
  - Eye movement patterns
- Conclusions

The Paper

- Published 1999

- Previously:

  - Reading/music reading
  - Steering a car
  - Copying block patterns

  Repetitive, not necessarily automatic

  - **Findings**: Eye movements extract information and are coupled with motor actions

- **This paper:** Studies eye fixation patterns during well-learned, automatic task

Fixations while picking up a mug

Experimental Setup

- Three subjects make a cup of tea while eye fixations are tracked
    - Kitchen seen once previously, objects/utensils moved around
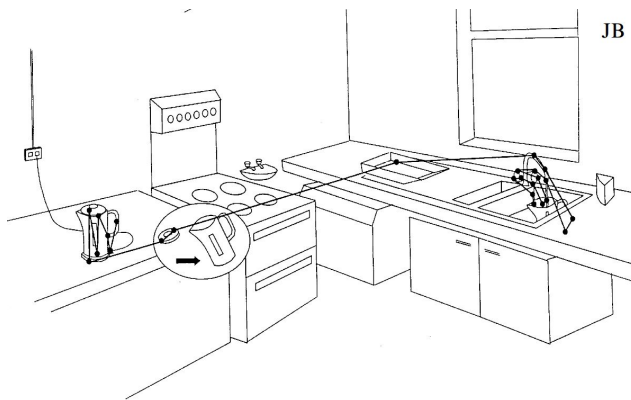

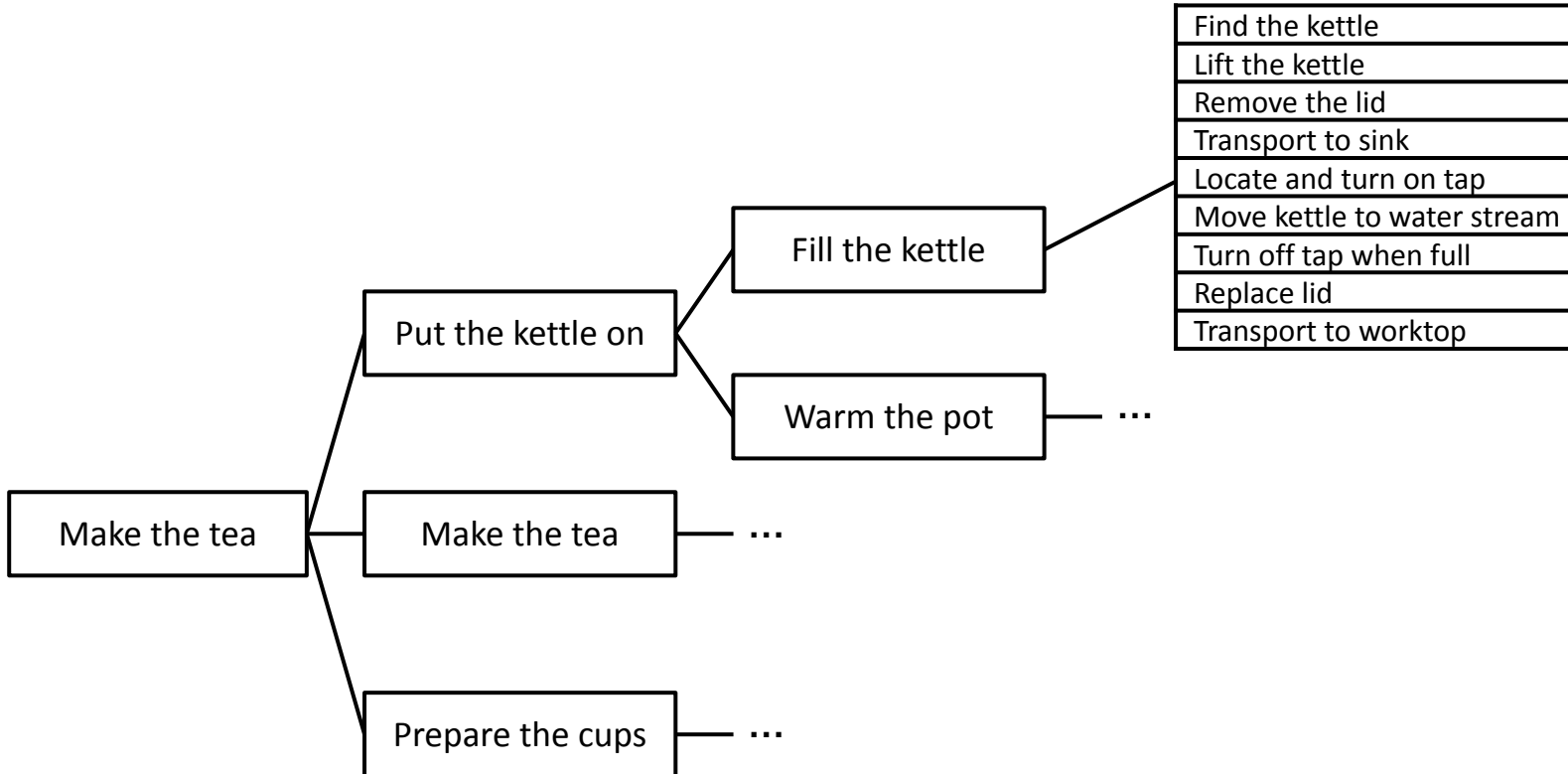Depiction of camera headgear

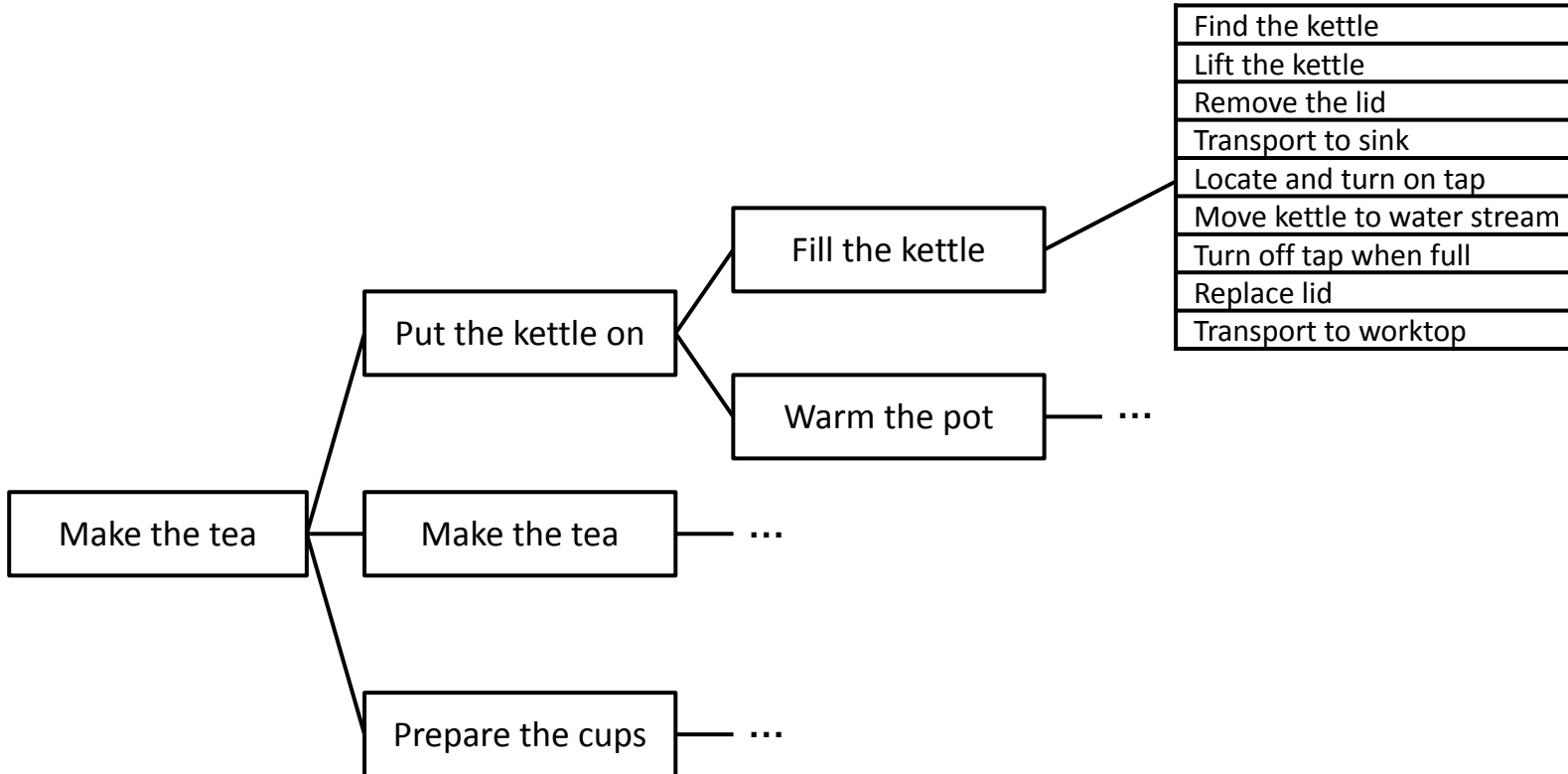Diagram of kitchen scene

The Tea-Making Task

The Tea-Making Task

| Find the kettle |
| Lift the kettle |
| Remove the lid |
| Transport to sink |
| Locate and turn on tap |
| Move kettle to water stream |
| Turn off tap when full |
| Replace lid |
| Transport to worktop |

Fill the kettle

Put the kettle on

Warm the pot ...

Make the tea

Make the tea ...

Prepare the cups ...

Eye movements

Analysis of Eye Movements in Tea-Making Task

Observation: Most fixations made during level-4 act are directed at object involved in act
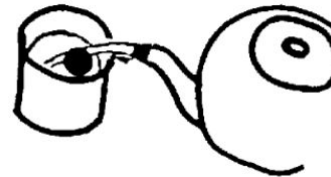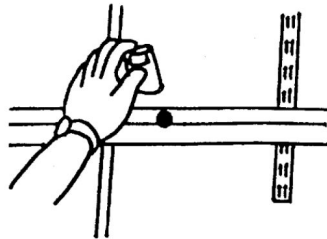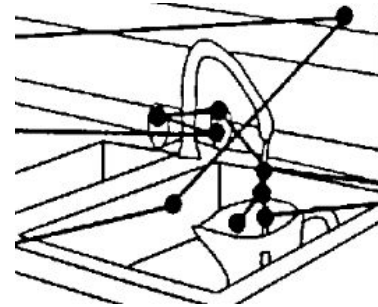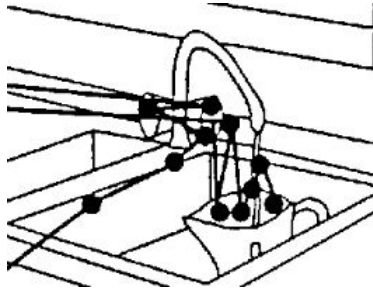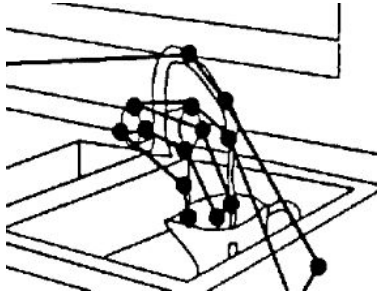
Fixation on objects precede interaction

- "Object-Related Actions" (ORAs) emerge
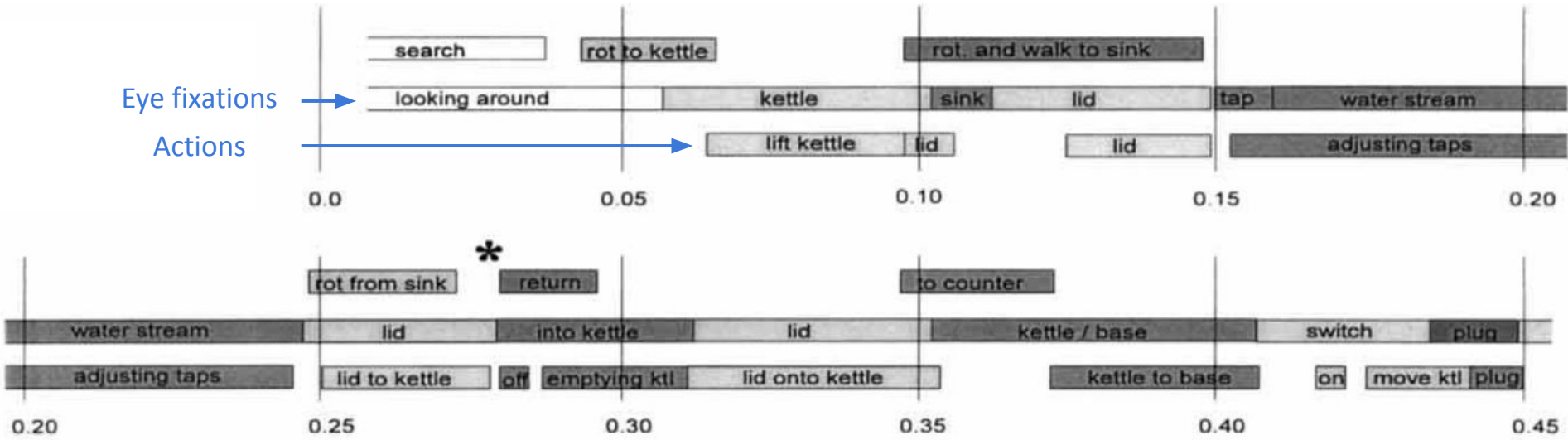
Fixation on objects precede interaction

- "Object-Related Actions" (ORAs) emerge
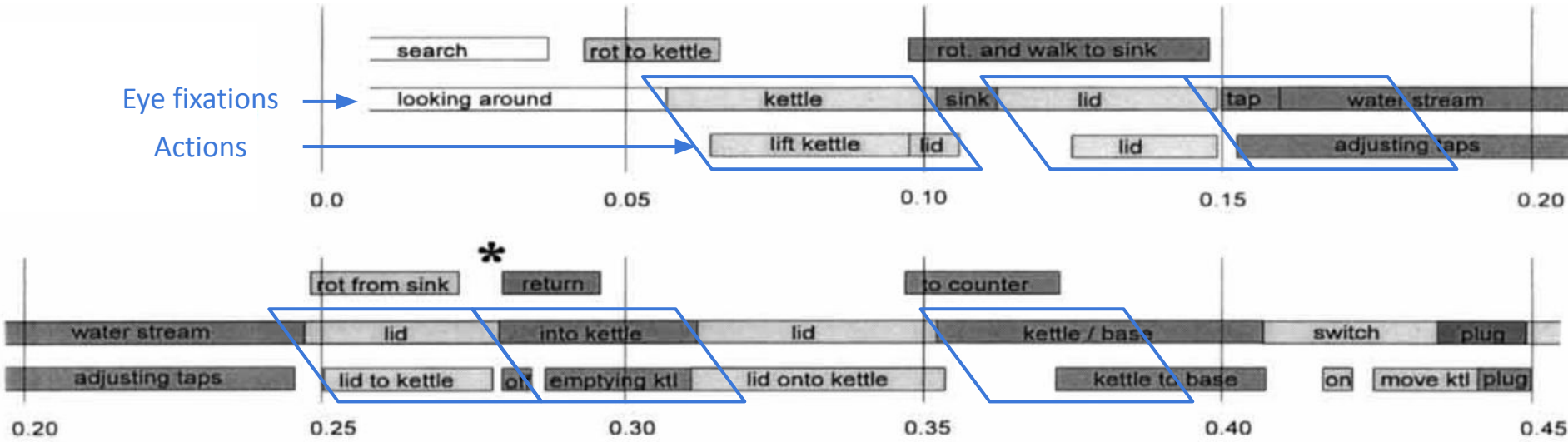
Time Relationship Across ORAs

- Trunk movement precedes first saccade to object by 0.61s
- Saccade precedes first manipulation by 0.56s
- Gaze moves on to next object 0.61s before last motor act completes
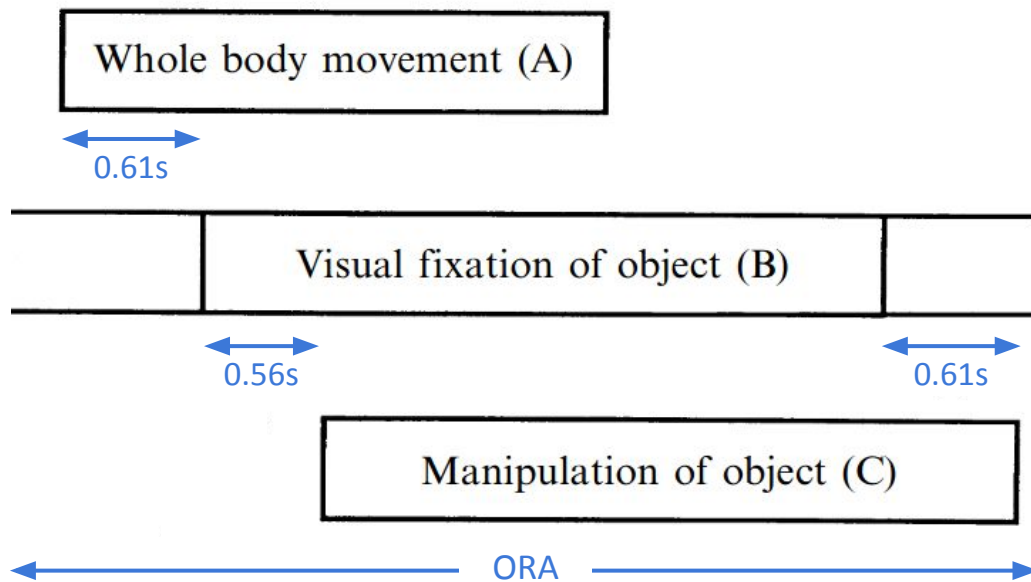
Categorization of fixations

- **Locate:** Fixate on object used later in action process
- **Direct:** Fixate on location or object that will be approached by hand
- **Guide:** Fixate between two objects about to interact
- **Check:** Fixate on location where state of variable is being assessed
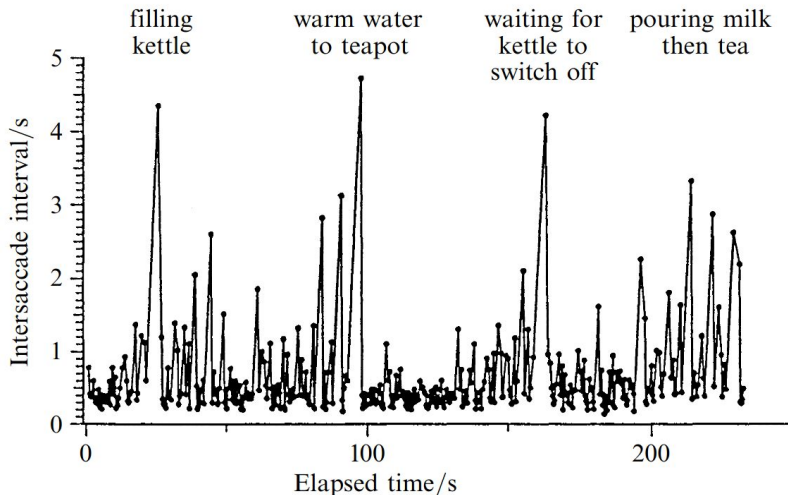
Checking actions have larger intersaccade intervals. Microsaccades ignored.
Rarely fixated:
- Hands
- Objects already in hand
- Familiar objects

Monitoring Role of the Eyes

- Tea-making = automatic task
  - Previously: Little to no conscious supervision
  - Supervised on low-level by eye fixation feedback
  - Closed-loop control just like other non-automatic tasks

Unconscious control   Conscious control

Natural Units of Action + Natural Time Scale

- Object-Related Actions (ORAs)
  - Link motor and sensory systems via a single object for ~3 seconds
  - Other studies find simple tasks have time scale ~2 seconds
    - Might imply natural task resolution that brain prefers
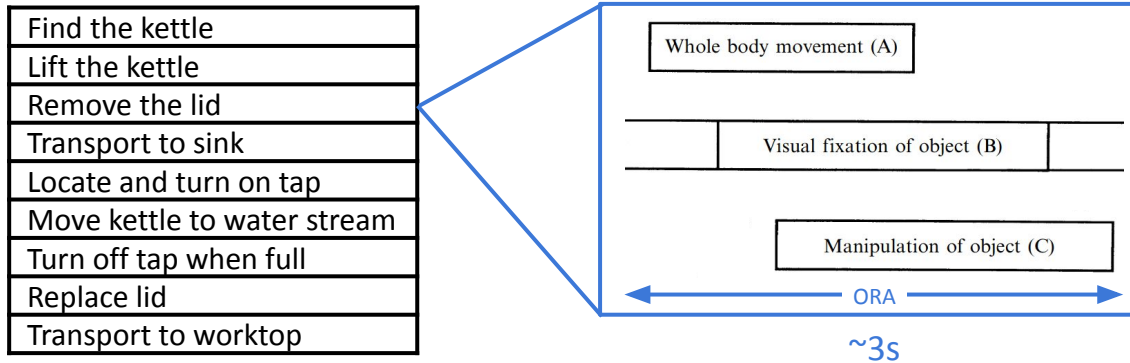
Find the kettle
Lift the kettle
Remove the lid
Transport to sink
Locate and turn on tap
Move kettle to water stream
Turn off tap when full
Replace lid
Transport to worktop

Whole body movement (A)
Visual fixation of object (B)
Manipulation of object (C)
←————— ORA —————→

~3s

What Directs Gaze?

- "Where one looks… seems to be driven principally by the retrieved memory 'script' for the activity"
  - Drawn to salience in "free-viewing", but cognition steps in when purposeful task is given
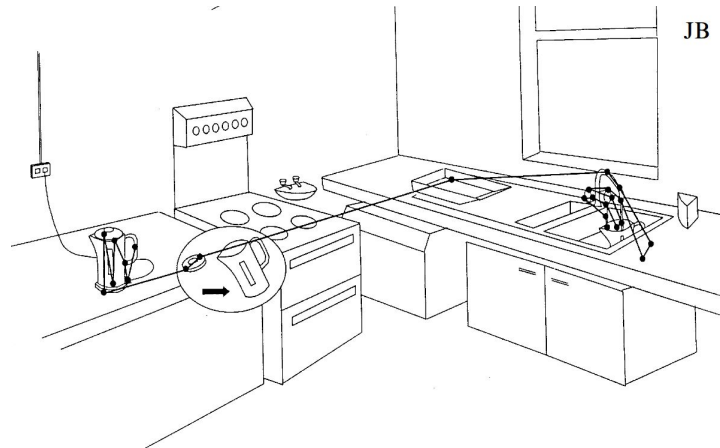
Fixations during task

Free-viewing fixations

What Directs Gaze?

- "Where one looks… seems to be driven principally by the retrieved memory 'script' for the activity"
  - Drawn to salience in "free-viewing", but cognition steps in when task is given
  - **Human allowance:** We can translate "script" (e.g. "add the sweetener") into semantically meaningful tasks
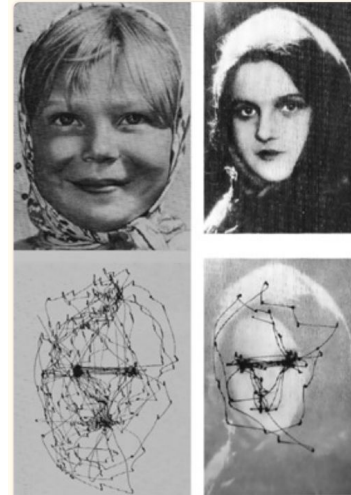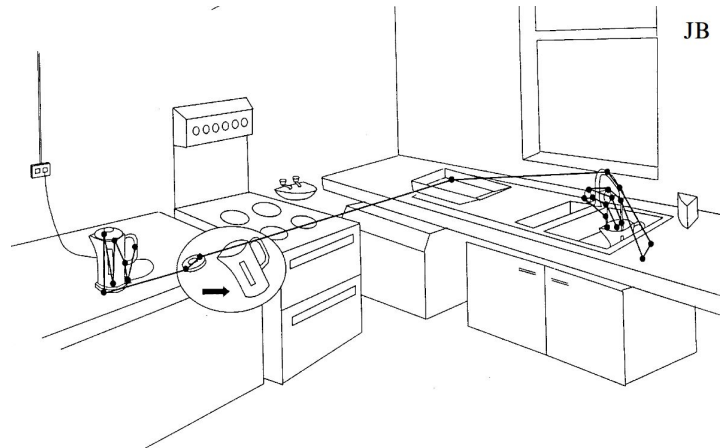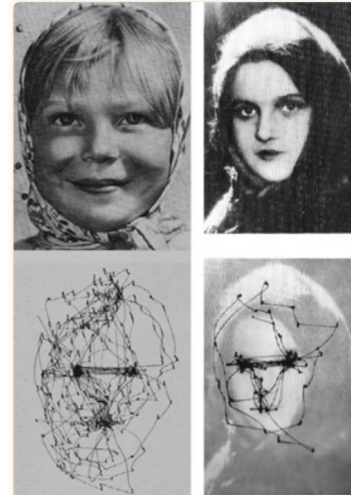

Fixations during task

Free-viewing fixations

Aside: Task extraction in large language models

GPT-3

Prompt

Robot: Hi there, I'm a robot operating in an office kitchen.
Robot: You can ask me to do various tasks and I'll tell you the sequence of actions I would do to accomplish your task.
Human: How would you put a water bottle and oatmeal next to the microwave?
Robot: 1. find a water bottle, 2. pick up the water bottle, 3. go to microwave, 4. put down the water bottle, 5. find an oatmeal, 6. pick up the oatmeal, 7. go to the microwave, 8. put down the oatmeal, 9. Done.

⋯

Later

Human: I have made my smoothie. What do I do with my dirty blender?
Robot: 1. find a sponge, 2. pick up the sponge, 3. go to the blender, 4. unplug the blender, 5. put the sponge into the blender, 6. add some soap, 7. turn on the blender, 8. wait for the blender to finish, 9. turn off the blender, 10. rinse the blender, 11. dry the blender, 12. plug in the blender, 13. done.

Level 3-4 Tasks

Unconscious Attention through eye movements

- Previously: "Conscious" attention supervises non-routine tasks
- Eye movements = unconscious attention?
  - Oculomotor systems attend to task, even if person does not

- **Conclusions:**
  - Automated activities still require continuous monitoring
    - Monitoring done through lower level mechanisms, unconsciously

Piazza Discussion: Determining Eye Fixation @59_f2

Eyes can extract information like luminance, and segmentation in a short time even during automated routine activities. How can eyes determine where to monitor by themselves (unconsciously)? Does it depend on an experience/script or environment?

- Goal oriented, prior task knowledge, internal mechanism with feedback signals?
- Paper explicitly mentions the saccade to transition to another object is based on prior knowledge. Idea of translating task script to gaze and then actions.
- In Class: Idea that conscious attention can actually harm automatic task performance. Ex: Conscious thinking about balancing an object can cause you to lose control.

Piazza Discussion: Further Experiments @59_f3, @59_f4

It would be interesting to see the same experiment performed on a more automated task (i.e. brushing teeth). It would also be nice to contrast these results with an experiment concerned with learning a new task.

- In the context of learning a new task or discovering affordances, do you believe eye movements still serve to monitor task completion?
- In Class: In the context of a new task, gaze may still be goal oriented or driven by curiosity and the desire to learn affordances.
- In Class: Gaze not only serves to monitor learning, but to simplify the task by gathering information.
- What would be the purpose of eye movements during a mindless task such as brushing?
- Mindless tasks could be explained with a salient-stochastic model, conscious attention may not apply in this case as well.

Piazza Discussion: "Normal Sight" @59_f7

When we have normal sight but have a temporary limitation (absence of light), do our eyes still work in a similar manner to guide actions, or is it replaced in part or completely by something else?

- Perhaps memory can remedy this, or maybe the brain can fill in gaps. Eye eventually adapts.
- **Followup:** How would people with a curable blindness adapt to vision in foreign situations?
- In Class: Vision is multimodal and the act of perception is a sensorimotor action. Motor actions and other sensory modalities can remedy this temporary limitation.

Piazza Discussion: Eyes in the Motor System @59_f1

If eye movement is not part of the motor system, we could say that fixation on object has monitoring functions. However, if we consider eye movement part of the motor system, isn't the guiding of eye movement also scripted?

- Interesting paradox that although the paper argues eye movement precedes motor act, the act of guiding the eyes may be a motor act itself. Should we consider eye movement a part of the motor system?
- Perhaps view eye movement as a monitoring script, and distinguish conscious and unconscious motor acts.

Piazza Discussion: Short Term Spatial Memory @59_f5

Subjects were only able to accurately redirect gaze to objects that were very recently manipulated. This supports the theory of efficient representations in human vision. Compared to computational models, humans do not benefit from memorizing exact copies of a scene.

- **Idea:** Representing world as constant voxels → voxel decay + rescanning
- In the context of robot perception, would it be more optimal to memorize the scene for future action? How would visual fixation play with robot perception?
- In Class: Attention and active updates are important in CNN architecture, we want to translate these to robot perception.

Piazza Discussion: Natural Setting @59f_12

I believe the authors wanted a natural setting but i'm not sure if the device in fig 1 maintains this - and I know there's a lot of other issues in behavior datasets where humans don't behave as they normally would since they know there being recorded. (I do understand this is more natural than previous settings).

- In Class: The backpack in the experiment seems unnatural and could impede behavior, but it was necessary as it housed the video recording device.
- In Class: The natural setting although unfamiliar would only affect the duration of locating objects. The fluidity when performing actions remains the same in the context of the automatic tea making task.

Piazza Discussion: Truly Automatic Task @59f_18

I thought that the idea that this was an "automatic" task in which the participants were thoughtlessly progressing through the task was a stretch. They were in a relatively unfamiliar kitchen under supervision while wearing a head-mounted camera as a volunteer for a study that had tasked them with making tea. It feels like their sole focus at the moment would be on making tea.

- The authors place a heavy emphasis on tea making being automatic to prove a point about closed loop control and monitoring. It would make more sense to keep object positions static. Perhaps have participants do another attentive activity while making tea.
- Paper explains it by saying that the novel environment will only affect the ease and speed of location objects and not the fluidity of action. Authors could have proved this claim.

Discussion Questions

- **Do you think vision is the primary sensory input that guides and monitors motor action? How would other sensory modalities play into the completion of routine or proactive tasks?**

- **Is "unconscious attention" a concept that applies to robots?; can a robot know how to unconsciously orient its sensory system to complete a routine task?**