COSC451: Artificial Intelligence

Lecture 8: Sequential structure in the experience of a reach-to-grasp action

Alistair Knott

Dept. of Computer Science, University of Otago

The aim for this lecture

What I want to persuade you of today:

- The process of executing a reach-to-grasp action has a characteristic sequential structure.
- The process of perceiving a reach-to-grasp action also has a characteristic sequential structure.
- Except for the first item, the sequences for execution and perception are basically the same.
- In each case, the agent and target of the action are each attended to *at two different times*.

Outline of the lecture

1. The decision to act or observe
2. Sequential structure in execution of a reach-to-grasp action
3. Sequential structure in perception of a reach-to-grasp action
4. The attentional states in action observation/execution
5. Summary

Outline of today's lecture

1. The decision to act or observe
2. Sequential structure in execution of a reach-to-grasp action
3. Sequential structure in perception of a reach-to-grasp action
4. The attentional states in action observation/execution
5. Summary

Definitions: an observer and an agent

We'll define someone called the observer, who either executes the reach-to-grasp action or perceives it.

> *Note: if he executes the action, he still observes it (from the perspective of the agent).*

We'll define the agent as the one who executes the action. Thus:

Action execution	Action perception
observer $=$ agent	observer $\neq$ agent

The decision to act or to observe

To begin with we envisage a context where the observer hasn't decided what to do.

- The observer can 'choose to act'.
- The observer can 'choose to perceive' something in the world.

My suggestion:

- Action execution: the initial action is a 'decision to act', which must happen before he can attend to an object in the world *as a target*.
- Action recognition: the initial action is an action of attention to an external object, which can *coincide* with a 'decision to observe'.

The decision to act

If the observer wants to act on an object, he has to attend to it *in a special motor sense*.

- He must enter a mode where perceptual stimuli are mapped to their motor affordances.
- In particular, points in a map of salient locations must be mapped to 'reach' movement vectors. (It's only then that a reach target can be selected.)

The decision to act

But surely the observer can attend to a cup *before* deciding to grab it?

- Noticing the cup could be what makes the observer *want* to reach for it.

I suggest:

- The observer notices the cup (in some neutral modality). This causes him to adopt a plan to grab the cup.
- The first step in this plan is to enter action execution mode. In this mode, his perceptual stimuli are mapped to movement vectors.
- He now 'selects a target object' in the motor modality. The cup easily wins the competition, and is selected as a target.

NB: the observer only selects the cup *as a target* after deciding to act.

The decision to observe

An observer enters 'perception mode' when he decides to monitor an external object or agent in the world.

- This decision temporarily recruits the observer's motor/intentional system, to allow representation of the motor actions of another agent.

I suggest the observer moves into perception mode when he attends to an 'interesting' object.

- 'Interesting' must mean something more than 'salient'.
  - Objects of particular types can be interesting.
  - Unusual objects can be interesting.
  - Moving or animate objects can also be interesting.

An asymmetry between deciding to act and to observe

‘Deciding to act’ is like *attending to onesself* as an ‘interesting’ object.

I suggest:

- ‘Deciding to act’ involves attending to onesself.
- ‘Deciding to observe’ involves attending to an external object.

Note: a ‘decision to observe’ shifts attention away from onesself.

- The agent of a ‘decision to act’ is attentionally identified.
- The agent of the ‘decision to observe’ is not attentionally represented.

Summary

The observer begins in a null context, where perceptual stimuli inform a decision to act or to observe.

- A 'decision to act' configures the observer's mirror system for action execution.
  The observer only selects an object as a target *after* having decided to act.
- A 'decision to observe' configures the observer's mirror system for action recognition.
  It is triggered by attention to an object. (Which will end up being the object the observation is 'about'.)

I'll now consider action execution and action recognition separately.

Outline of today's lecture

1. The decision to act or observe
2. Sequential structure in execution of a reach-to-grasp action
3. Sequential structure in perception of a reach-to-grasp action
4. The attentional states in action observation/execution
5. Summary

Choosing a target and choosing a motor programme

When the observer has decided to act, he has two further decisions:

- What object(s) is he going to act *on*?
- What motor programme is he going to execute?

Note: deciding on an object *as a target* means choosing a movement vector in the ‘reach’ motor pathway. Choosing a *motor programme* involves more than this: we need to choose a hand trajectory and a finger preshape sequence.

I suggest: he has to choose a target before choosing a motor programme.

Choosing a target > choosing a motor programme

Some arguments:

- Agents typically attend to a target they reach for very early in the reach trajectory (see e.g. Johansson *et al.*, 2001)
- Selecting a reach target triggers a shift of visual attention to this target (see e.g. Schneider & Deubel, 2002). The link probably involves F7v, the supplementary eye fields).
- Many computations in the 'grasp' pathway presuppose that the target is attended to. (E.g. computing the shape of the object requires attention to the object.)
- In macaque, many 'canonical neurons' in F5 only activate if the monkey fixates an object requiring the associated grasp (see e.g. Gallese *et al.*, 1996).

Outline of today's lecture

Identifying a target and classifying the action

When the observer has decided to observe, he must determine two things:

- What object(s) the observed agent is acting on (if any)
- What the observed agent's motor programme is (if anything)

I suggest: if the observed action is a reach-to-grasp, the observer identifies the target of the action before classifying the action.

Attention to the target > classifying the action

Some arguments:

- Observers of a reach-to-grasp action saccade to the intended target *early* in action monitoring (Flanagan and Johansson, 2003).
- Many mirror neurons don't fire if their associated action is 'pantomimed', without a target object (Gallese *et al.*, 1996).
- Computational models of action recognition tend to assume that observers represent the trajectory of the hand *in relation to the target*. (See e.g. Oztop and Arbib, 2002).
- It's hard to see how the mirror system can be trained unless the observer of an action establishes joint attention with the agent.

Some more supporting data

Attention to agent $>$ attention to target:

- Webb *et al.* (in preparation): observers reliably saccade to the agent of an observed reach-to-grasp action, and then to the target. (Again, the target saccade is anticipatory.)

Attention to agent $>$ classifying the action:

- Nelissen *et al.* (2005): most F5 mirror neurons only respond to an action if the agent of the action is in view. (They don't respond to 'disembodied arms'.)

Summary

Action execution:

- Attention to agent $>$ attention to target (as a target)
- Attention to target $>$ activation of motor programme

Action observation:

- Attention to agent $>$ attention to target
- Attention to target $>$ activation of motor programme

Summary

If an observer *O* executes the action of grabbing a cup:

- *O* attends to himself.
- *O* attends to the cup (as a target).
- *O* activates the ‘grab’ motor programme.

If *O* perceives an external agent *A* grabbing a cup:

- *O* attends to *A*.
- *O* identifies *A*’s intention, and attends to the cup.
- *O* activates the ‘grab’ motor programme.

Some objections: action observation

Does observing a reach-to-grasp action *have* to involve that sequence?

- Surely I can recognise an action even if I don't see the agent?
- Surely I can look at the target first, and then look at the agent?
- Surely I can recognise an action from a still picture?

Some objections: action observation

Does observing a reach-to-grasp action *have* to involve that sequence?

- Surely I can recognise an action even if I don't see the agent?
  Then we would report it as a *passive sentence*.
- Surely I can look at the target first, and then look at the agent?
- Surely I can recognise an action from a still picture?

Some objections: action observation

Does observing a reach-to-grasp action *have* to involve that sequence?

- Surely I can recognise an action even if I don't see the agent?
  Then we would report it as a *passive sentence*.
- Surely I can look at the target first, and then look at the agent?
  We don't know how that is reported yet.
- Surely I can recognise an action from a still picture?

Some objections: action observation

Does observing a reach-to-grasp action *have* to involve that sequence?

- Surely I can recognise an action even if I don't see the agent?
  Then we would report it as a *passive sentence*.
- Surely I can look at the target first, and then look at the agent?
  We don't know how that is reported yet.
- Surely I can recognise an action from a still picture?
  We probably still recognise the intention of the agent. And that has a sequential structure.

Some objections: action execution

Does executing a reach-to-grasp action *have* to involve that sequence?

- Surely I can reach for an object without looking for it?
- Surely I can decide what action I'm going to do long before I attend to a target object?

Some objections: action execution

Does executing a reach-to-grasp action *have* to involve that sequence?

- Surely I can reach for an object without looking for it?
  You are probably using some form of *covert attention* in this case.
- Surely I can decide what action I'm going to do long before I attend to a target object?

Some objections: action execution

Does executing a reach-to-grasp action *have* to involve that sequence?

- Surely I can reach for an object without looking for it?
  You are probably using some form of *covert attention* in this case.
- Surely I can decide what action I'm going to do long before I attend to a target object?
  The 'intention to grab a cup' is probably an intention to do a sequence of things: first find a cup, then grab it.
  Also, the *specific grasp action* can only be selected when you know what you're grasping for.

Some objections: action observation/execution modes

Surely I can be watching an action at the same time I'm doing one?
E.g. playing tennis?

Some objections: action observation/execution modes

Surely I can be watching an action at the same time I'm doing one? E.g. playing tennis?

- Doing an action while watching another one is *dual task performance*.
  Some dual tasks are performed by switching rapidly between tasks.

Some objections: action observation/execution modes

Surely I can be watching an action at the same time I'm doing one? E.g. playing tennis?

- Doing an action while watching another one is *dual task performance*.
  Some dual tasks are performed by switching rapidly between tasks.
- It may also be that such activities involve creating a single 'joint agent'.
  C.f. *John and Mary played tennis*...

Outline of today's lecture

Actions and their reafferent feedback

A SM action signal should be distinguished from the perceptual signals which result from its execution.

For attentional actions:

- An action signal is a direction of attention;
- Reafferent feedback is a representation of the attended object.

For motor actions:

- An action signal is the activation of a motor programme in (pre)motor cortex;
- Reafferent feedback might include a representation of the agent *as an agent*.

The role of reach-to-grasp actions in creating crossmodal object representations

We need to represent an agent *as an object* and *as an agent*.

We need to represent a target *as an object* and *as a set of motor affordances*.

- A reach-to-grasp action involves attending to the agent in both ways. It provides an opportunity to learn the associations between them.
- Ditto the target.

The 'stable grasp' state

If the observer successfully performs a reach-to-grasp action, he achieves a stable grasp state. This is particularly useful for learning a cross-modal representation of the target.

- The observer can learn to map a *visual* representation of the location of the grasped object to the current *motor* representation of the position of his hand.
- The observer can learn to map a *visual* representation of the shape of the grasped object to the current *motor* representation of the shape of his hand.

Static and dynamic agent representations

An agent is an ‘object’: recognisable by a characteristic pattern.

An agent is also a dynamic entity:

- An articulated entity, which moves in certain characteristic ways.
- An entity associated with particular dispositions to act. (Not all agents act the same way in the same circumstances!)

An observer needs to relate these concepts together, so when he attends to an agent he can predict how s/he will act.

Outline of today's lecture

1. The decision to act or observe
2. Sequential structure in execution of a reach-to-grasp action
3. Sequential structure in perception of a reach-to-grasp action
4. The attentional states in action observation/execution
5. Summary

The sequence for perception of a reach-to-grasp

The sequence for perception of a reach-to-grasp

Operation 1	*O* attends to an external agent, configuring his mirror system circuit for action perception.

The sequence for perception of a reach-to-grasp

Operation 1	*O* attends to an external agent, configuring his mirror system circuit for action perception.
State 1	*O* receives reafferent feedback from this operation; the percept 'man'.

The sequence for perception of a reach-to-grasp

Operation 1	*O* attends to an external agent, configuring his mirror system circuit for action perception.
State 1	*O* receives reafferent feedback from this operation; the percept 'man'.
Operation 2	*O* establishes joint attention with the agent, and attends to another object (the cup).

The sequence for perception of a reach-to-grasp

Operation 1	*O* attends to an external agent, configuring his mirror system circuit for action perception.
State 1	*O* receives reafferent feedback from this operation; the percept 'man'.
Operation 2	*O* establishes joint attention with the agent, and attends to another object (the cup).
State 2	*O* receives feedback from this operation: the percept 'cup'.

The sequence for perception of a reach-to-grasp

Operation 1	*O* attends to an external agent, configuring his mirror system circuit for action perception.
State 1	*O* receives reafferent feedback from this operation; the percept ‘man’.
Operation 2	*O* establishes joint attention with the agent, and attends to another object (the cup).
State 2	*O* receives feedback from this operation: the percept ‘cup’.
Operation 3	*O* initiates a process of biological motion classification, which results in the action ‘grab’ being activated in *O*’s premotor cortex.

The sequence for perception of a reach-to-grasp

Operation 1	*O* attends to an external agent, configuring his mirror system circuit for action perception.
State 1	*O* receives reafferent feedback from this operation; the percept 'man'.
Operation 2	*O* establishes joint attention with the agent, and attends to another object (the cup).
State 2	*O* receives feedback from this operation: the percept 'cup'.
Operation 3	*O* initiates a process of biological motion classification, which results in the action 'grab' being activated in *O*'s premotor cortex.
State 3	As a corollary of this process, *O* re-attends to the agent *as an agent*.

The sequence for perception of a reach-to-grasp

Operation 1	*O* attends to an external agent, configuring his mirror system circuit for action perception.
State 1	*O* receives reafferent feedback from this operation; the percept ‘man’.
Operation 2	*O* establishes joint attention with the agent, and attends to another object (the cup).
State 2	*O* receives feedback from this operation: the percept ‘cup’.
Operation 3	*O* initiates a process of biological motion classification, which results in the action ‘grab’ being activated in *O*’s premotor cortex.
State 3	As a corollary of this process, *O* re-attends to the agent *as an agent*.
State 4	*O* re-attends to the cup, in the course of perceiving the agent establishing a stable grasp.

The sequence for execution of a reach-to-grasp

The sequence for execution of a reach-to-grasp

Operation 1	*A* attends to himself, configuring his mirror system circuit for action execution.

The sequence for execution of a reach-to-grasp

Operation 1	*A* attends to himself, configuring his mirror system circuit for action execution.
State 1	*A* receives reafferent feedback that this operation succeeded.

The sequence for execution of a reach-to-grasp

Operation 1	*A* attends to himself, configuring his mirror system circuit for action execution.
State 1	*A* receives reafferent feedback that this operation succeeded.
Operation 2	*A* selects an object to reach for (the cup), and hence executes an action of attention to the cup.

The sequence for execution of a reach-to-grasp

Operation 1	*A* attends to himself, configuring his mirror system circuit for action execution.
State 1	*A* receives reafferent feedback that this operation succeeded.
Operation 2	*A* selects an object to reach for (the cup), and hence executes an action of attention to the cup.
State 2	*A* receives feedback from this operation: the percept 'cup'.

The sequence for execution of a reach-to-grasp

Operation 1	*A* attends to himself, configuring his mirror system circuit for action execution.
State 1	*A* receives reafferent feedback that this operation succeeded.
Operation 2	*A* selects an object to reach for (the cup), and hence executes an action of attention to the cup.
State 2	*A* receives feedback from this operation: the percept ‘cup’.
Operation 3	*A* selects an action category (‘grab’) and begins to execute the grab action.

The sequence for execution of a reach-to-grasp

Operation 1	*A* attends to himself, configuring his mirror system circuit for action execution.
State 1	*A* receives reafferent feedback that this operation succeeded.
Operation 2	*A* selects an object to reach for (the cup), and hence executes an action of attention to the cup.
State 2	*A* receives feedback from this operation: the percept 'cup'.
Operation 3	*A* selects an action category ('grab') and begins to execute the grab action.
State 3	*A* receives feedback from the process of action execution, which includes a percept of himself 'as an agent'.

The sequence for execution of a reach-to-grasp

Operation 1	*A* attends to himself, configuring his mirror system circuit for action execution.
State 1	*A* receives reafferent feedback that this operation succeeded.
Operation 2	*A* selects an object to reach for (the cup), and hence executes an action of attention to the cup.
State 2	*A* receives feedback from this operation: the percept 'cup'.
Operation 3	*A* selects an action category ('grab') and begins to execute the grab action.
State 3	*A* receives feedback from the process of action execution, which includes a percept of himself 'as an agent'.
State 4	*A* re-establishes the cup in the haptic modality.

SM signals during experience of a reach-to-grasp

	Transient signals	
	Action signals	Reafferent signals
	attend_agent	*attending_to_agent*
	attend_cup	*attending_to_cup*
	grasp	*attending_to_agent*
		attending_to_cup

A framework for action representation: deictic routines

Ballard *et al.* (1997) noticed that many cognitive processes are *inherently sequential*.

- The representations in ‘perceptual’ neural areas are *transitory*: they reflect the agent’s current focus of attention.
- These representations often drive a *change* in the agent’s attention.
- A deictic routine is a sequence transitory neural representations linked by attentional actions $R_1, A_1, \ldots, R_n, A_n, R_{n+1}$.
  - R_i enables A_i. A_i brings about R_{i+1}.
- To express the ‘meaning’ of a transitory neural representation, we must often make reference to the deictic routine which brought it about.