# Causative motor actions and the causative alternation

Jeremy Lee-Hand and Alistair Knott*

Dept of Computer Science, University of Otago
P.O. Box 56, Dunedin, New Zealand

Email: leehandjeremy@gmail.com; alik@cs.otago.ac.nz;

*Corresponding author

## Abstract

Certain verbs (e.g. *bend*, *break*) can undergo the **causative alternation**, appearing both in a transitive form (*John bent the lever*) and an intransitive form (*The lever bent*). To explain this alternation, linguists have proposed that the transitive form of the alternation conveys an action with an explicitly causative structure—for example that *John bent the lever* indirectly conveys the proposition *John caused [the lever bent]*. While this analysis is syntactically appealing, it makes strong assumptions about explicit representations of causality in the motor system that must be independently justified. In fact there is a large body of research within motor neuroscience showing that motor programmes can be defined by the perceptual effects they cause. In this paper we argue that the causative action representations posited by linguists are direct references to the effect-based action representations identified by motor neuroscientists. We develop a neural network model of how an agent can learn actions that are defined by the effects they bring about, and deploy this model in an account of the syntax of causative sentences.

**Keywords:** causative alternation, unaccusative verbs, causative actions, event codes, common coding, neural networks

## Acknowledgements

We thank Barbara Kaup for helpful discussions, and Tim Neumegen for developing the GraspProject software.

# 1 Causative action representations in language and the motor system

## 1.1 A syntactic model of the causative alternation

This paper develops a model of the syntactic phenomenon called the **causative alternation**. Verbs that undergo this alternation can be used both transitively and intransitively, as illustrated in Examples 1 and 2.

(1) John bent the lever.

(2) The lever bent.

The interesting fact about this alternation is that when the verb is used transitively, the entity undergoing the described action (the lever in the above examples) appears as the object of the verb, while when the verb is used intransitively, this same entity appears as the subject. Not all transitive verbs undergo the causative alternation. An example of a verb that does not is *grab*, as Examples 3 and 4 show.

(3) John grabbed the lever.

(4) ? The lever grabbed.

How can we account for the ability of verbs like *bend* to undergo the causative alternation? One possibility is that such verbs are systematically ambiguous. On this hypothesis there are two different senses of the verb *bend*: one in which the object being bent appears as the subject, the other in which it appears as the object. But this explanation is uneconomical—it requires us to stipulate the semantic roles played by the arguments of these two verbs individually, and prevents us telling a general story about the semantic roles played by subjects and objects. A proposal that is more appealing to linguists is that the transitive sentence *John bent the lever* actually conveys a more complex proposition, namely that John *caused the lever to bend* (see e.g. Dowty, 1979; Parsons, 1990; Levin and Rappaport Hovav, 1995). This proposal assumes a syntactic theory in which sentences have an 'underlying' syntactic or semantic structure distinct from their surface form—an idea associated with Chomskyan accounts of syntax (see e.g. Chomsky, 1995). The basic suggestion is that the underlying structure of Example 1 is *John caused [the lever bent]*, as shown schematically on the left of Figure 1. In this structure, there are two clauses, representing two events: one clause (whose subject is


Figure 1: Derivation of *John bent the lever* by movement from an underlying syntactic structure

*John*) reports an action that causes something to occur; the other clause (whose subject is *the lever*) reports

the event which is caused to occur. In this representation, *the lever* stands in the same syntactic relation to the verb *bent* as it does in the intransitive Example 2. The surface structure of Example 1 is produced by moving the lower verb *bent* into the position of the higher verb *caused*, as shown on the right of Figure 1. This typically happens through operations that are independently motivated in the grammar—for instance, the Chomskyan operation of head raising (see Section 6). The proposal accounts for sentences like Example 3 by suggesting that these sentences have an underlying structure that is similar to their surface structure, featuring just a single clause: for this reason they do not undergo the causative alternation.

There is some empirical support for the idea that verbs undergoing the causative alternation have a more complex structure than those that do not. For instance, McKoon and Love (2011) showed that verbs like *bend* take longer to process in a lexical decision task than verbs like *grab*, when other factors known to influence processing times are controlled for, and that sentences like *John bent the lever* take longer to process than sentences like *John grabbed the lever*.

The proposal that transitive sentences featuring verbs like *bend* have an explicitly causative underlying structure is quite widely accepted by linguists. However, it makes a strong assumption about the nature of cognitive action representations: namely that actions in the motor system can be represented as operations that cause certain states or events, rather than simply as patterns of motor movement. To corroborate the linguists' proposal, it is important to examine whether motor actions can indeed be represented by the effects they bring about—and if so, how this happens. This is not straightforward, because the relations between a motor action and the effects it produces are complex and action-dependent. For instance, the temporal relationship between an action and its effects is quite variable. In some cases the caused movements can go on for longer than the causative action: for instance the motor action that causes a ball to roll may terminate long before the ball finishes rolling. In other cases, such as folding or scrumpling a piece of paper, the causative motor movement occurs simultaneously with the effect it produces. At the same time, a causative action often begins well before the agent even makes contact with the object in which the effect is to be produced. To bring about a particular change of state in a target object, the hand must often take a particular trajectory towards the target: for instance to squash an object, the hand must often approach the object from above; to lift an object, the hand typically approaches the object with a trajectory dependent on the object's weight (see Roy *et al.*, 2013). The representations of causative actions that feature in syntacticians' accounts of the causative alternation are discrete and symbolic: to justify them as cognitive representations, an account must be given of how they reflect the complexity of the actual relationship between causative motor actions and the effects they produce. Whether this can be done is currently an open question.

### 1.2 Causative actions in motor neuroscience

Independently of research in linguistics, there is a large body of research in neuroscience indicating that motor actions can be represented by the sensory effects they produce. The idea has a long history, but in recent research it is most strongly associated with Prinz's (1997) theory of 'common coding' and Hommel *et al.*'s theory of of 'event codes' (Hommel, Müsseler, Aschersleben, and Prinz, 2001). The key idea uniting these models is that motor programs are not defined purely within the motor domain: their neural representation includes a representation of the effects they are expected to have on the world, as apprehended by the perceptual system. This position can be supported both on theoretical grounds and through experiments; we will give brief examples of each kind of argument.

Theoretically, a strong argument for this view of action representation comes from considerations about how actions are learned. It is uncontroversial that an agent's repertoire of motor programs is learned through some kind of reinforcement. A reinforcing signal is a sensory signal. When an agent executes a motor program and generates a rewarding signal, an association is made between the sensory signal and this particular program. After a certain amount of training, if all goes well, the sensory signal will become associated with a range of related motor movements, which bring it about in different ways or under different circumstances, perhaps in ways which are parameterised or organised by features of the sensory stimulus. At this point, if the agent activates the sensory signal, this will bring about one of these movements, and result in reward. But equally importantly, the group of motor movements associated with the sensory signal can now be thought of as comprising an *action category*, in virtue of their shared ability to evoke this signal. Categories are defined around central concepts or prototypes, and in this case the unifying concept is a sensory one. For this reason, it makes sense to talk about action categories as being defined by the sensory effects they bring about.

Experimentally, the idea that actions are defined by their effects has been supported in several ways. For instance, there have been many studies exploring variations on the well-known stimulus-response compatibility effect (Simon, 1969). A good example is a study by Hommel (1993). Here subjects had to respond to an auditory stimulus by pressing a button, either with the left or right hand. The tone of the auditory stimulus indicated which button the subject should press. But as a distracting factor, the stimulus was also presented either on the left or the right. The classical stimulus-response compatibility effect is that subjects are slower to respond if the spatial location of the stimulus is incompatible with the hand which must respond. In Hommel's experiment, button presses generated a reafferent visual stimulus whose location could be decoupled from the location of the hand pressing the button, to explore whether the compatibility effect operates

in the domain of motor movements or that of their sensory consequences. Button presses consistently produced a visual stimulus: illumination of a light. In one condition the light appeared on the same side as the hand (e.g. left button presses illuminated a light on the left), while in another it appeared on the opposite side (e.g. left button presses illuminated a light on the right). Hommel found that the stimulus-response compatibility effect depended on compatibility with the perceptual effects of button-presses, rather than on the hand which was used. This shows that the way subjects encode actions does make some reference to their sensory consequences—at least enough to interfere with stimulus-response mappings. Effect-based representations of motor actions are also supported by several studies of the neural representation of actions; see for instance Umiltà *et al.* (2008); Matsumoto *et al.* (2003); Moore and Obhi (2012).

While there is good empirical evidence that actions are linked to their perceived effects, it is still far from clear how this link is represented in the brain. Many studies have shown an association between motor actions and their perceived effects, but the neural mechanisms that *establish* this association have not yet been clearly identified. These mechanisms not only establish a relation between actions and perceived effects, but somehow label this relation as one of causation, rather than (for instance) simply of co-occurrence or or temporal succession. So there are also many open questions in neuroscientific accounts of causative actions.

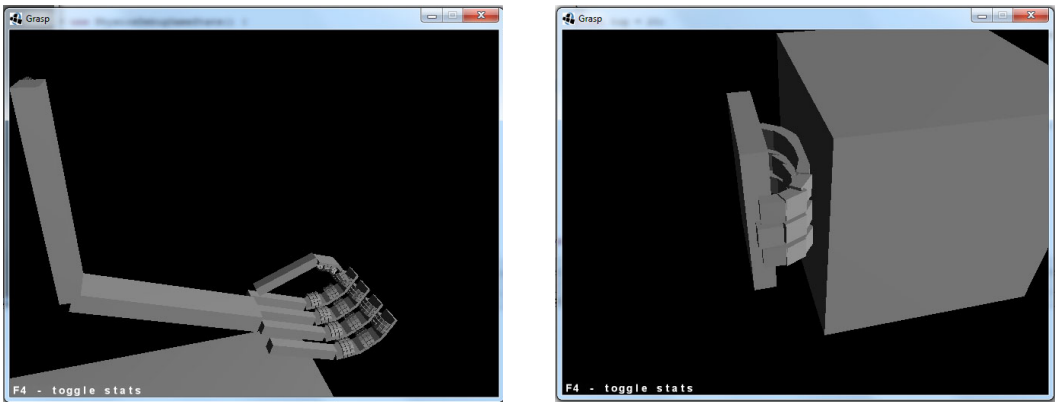
### 1.3 A motor model of the causative alternation

Our goal in the current paper is to connect the representations of causative actions posited by linguists with models of causative actions developed by motor neuroscientists. We will argue that the symbolic 'cause' predicate featuring in the underlying syntactic structure of sentences like *John bent the cup* references something specific in the motor circuitry that implements the action of bending. We express this proposal by developing a computational model of the learning and control of causative actions. The model is partly intended as a proposal about the neural mechanisms that associate motor actions with perceived effects, to account for the effect-based action representations that have been found in motor neuroscience. But we also suggest the model may be able to contribute to an acccount of causative constructions in language. An interesting recent suggestion in cognitive science is that high-level cognitive representations, including linguistic representations, directly recruit or evoke structures in the sensorimotor system (see e.g. Barsalou 2008, Glenberg and Gallese, 2012). Adopting this 'embodied' conception of linguistic representations, we propose that the underlying structures posited by linguists to account for the causative alternation (e.g. 'John caused [the lever bent]') make direct reference to the effect-based representations of actions found in the motor system. We introduce the computational model of causative actions in Sections 2–5. We discuss

how the model may serve as the basis for an account of the causative alternation in Section 6.

## 2 A platform for learning and control of simulated actions

Our computational model is implemented in a software environment for simulating hand/arm actions called GraspProject (Lee-Hand *et al.*, 2012; for details see Neumegen, 2013). The environment is built on top of the JMonkey games engine, which uses the Bullet physics engine to define objects made up of linked rigid bodies, and OpenGL to render graphical views of these. GraspProject provides a simple model of the hand and arm, with three degrees of freedom in the arm (two at the shoulder and one at the elbow) and one in the hand controlling grip aperture (see Figure 2a). It also provides a fairly rich model of the touch sensors




(a) (b)

Figure 2: (a) The hand/arm model used in simulations. (b) Detail of a single finger pad

in the fingers. Finger pads are modelled as deformable grids of rigid bodies connected by springs (see Figure 2b). Information about light touches is provided by collision detectors on each pad, and information about stronger touches which deform the surface of the skin is read from the joint angles between adjacent pads.

## 3 Architecture of the motor control network

Our model of the motor system is a neural network for learning hand actions directed at target objects. It provides a simple model of some aspects of infant motor development.

The general architecture of the network is shown in Figure 3. It consists of three sub-networks arranged in sequence. These are assumed to be trained at three successive developmental stages, by reward signals of different degrees of complexity. In this scheme, the system is initially rewarded by very simple sensory signals, which train a simple motor circuit, but as learning takes place in this circuit, more complex reward

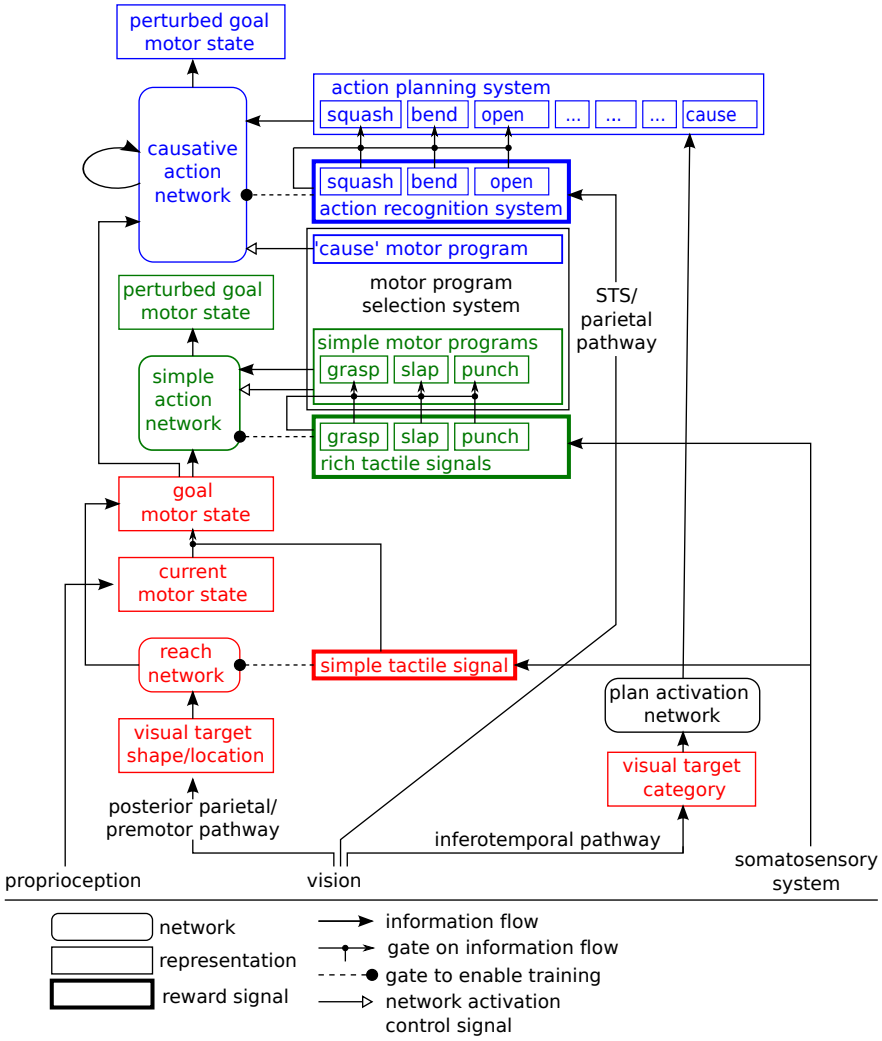



Figure 3: Architecture of the motor control network

signals become available, which in turn train higher-order circuits. In this section we will give an overview of the networks; their training and evaluation are described in Section 4.

The first network to be trained is called the **reach network** (see the red part of Figure 3, and Figure 4). This network learns a function which maps a visual representation of the location of the target object onto a goal motor state of the hand and arm. During training, the agent visually attends to objects in its perispace, and executes hand/arm actions at random. Sometimes these actions result in its hand touching the target, evoking a touch signal (the simple touch signal). This signal is intrinsically rewarding (as in Oztop *et al.* 2004). The touch signal has two functions. First, it allows a proprioceptive representation of the agent's current motor state to be copied into the medium holding its goal motor state (see the gating link

terminating on the connection between the current and goal motor states in Figure 3). Second, it allows the reach network to be trained, so that the current visual representation of the target object is associated with this newly specified goal motor state, and similar presentations of the target in the future will automatically elicit an appropriate motor goal (see the gating link terminating on the reach network).

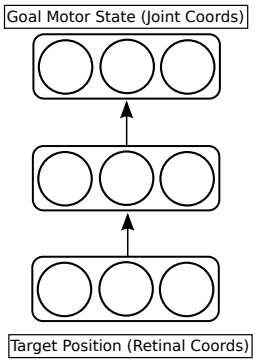


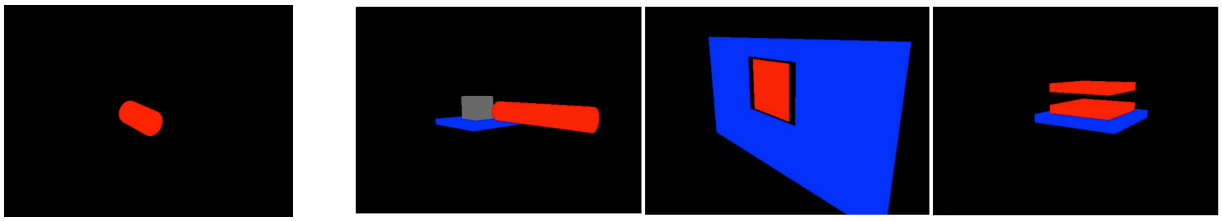

Figure 4: Detailed architecture of the reach network

This simple circuit implements a particular version of Hommel *et al.*'s model of event codes. Learning in the circuit creates what can be thought of as a single simple action category, associated with the sensory representation of a touch to the hand: after training, when the reach network is presented with the visual location of a target object, it will activate a motor goal which when achieved will reliably elicit this sensory representation. Motor goals in the circuit are associated with sensory stimuli in three ways. Any representation in the motor goal medium is implicitly associated with one particular reward stimulus (a simple touch sensation). Specific motor goals are associated axiomatically with specific motor states (sensed proprioceptively) when the reward stimulus is evoked. And specific motor goals are also associated through learning with arbitrary sensory stimuli (in this case visual), which carry information about the motor states associated with reward signals. Again this happens at the time the reward stimulus is evoked. The key devices in the circuit are reward-gated copy and learning operations. These devices are replicated in the other two networks.

The reach network generates a motor goal—but of course there must also be a mechanism which achieves this goal. In our current model, we assume this mechanism is a simple feedback motor controller. This controller takes the current motor state and the goal motor state and generates a motor signal proportional to the difference between them, in a direction which reduces this difference. (The controller is not shown in Figure 3.) A feedback controller does not need to be trained; it is a simple circuit, which is present innately in

many motor systems (see e.g. Kawato *et al.* 1987). (We use a PID controller; see e.g. Araki 2006). However, mature motor control involves a mixture of hardwired feedback control and learned *feedforward* control (see again Kawato *et al.*, 1987). Feedforward control exploits learning about the properties of the agent's motor system to optimise action trajectories. If we think of the feedforward controller in sufficiently general terms, we can say that it is through learning in this controller that an agent can acquire a repertoire of different action categories. Different actions (like grabbing or punching or slapping) have different characteristic trajectories of the hand and fingers; the feedforward control system somehow learns about the distinct effects of particular trajectories and creates action categories associated with each. However, it is not clear how different trajectories are represented in the biological motor control system. There is good evidence that agents do not compute detailed trajectories in advance; these are only generated 'on the fly', as an action is actually underway (see e.g. Cisek, 2005). Our network implements a particular idea about how trajectories are represented. We assume that the agent evoking a goal motor state can generate learned *perturbations* of this goal state as an action is under way, which deviate the hand from the normal course it would take under simple feedback control. For instance, to generate a trajectory bringing the hand onto the target from above, the goal state could be temporarily perturbed to a point above the target, so the hand initially moves higher than it would normally do. This idea is discussed in more detail and evaluated in Lee-Hand *et al.* (2012). This kind of learning takes place in the second network in our model, the **simple action network** (see the green part of Figure 3, and Figure 6).

The simple action network learns to map a goal motor state onto a perturbed goal motor state, which is applied at the start of a reach action and removed when the hand is at a specified distance from the target. This network becomes engaged when the reach network (which provides its input) reliably generates actions that lead to reward signals. Our proposal is that simple action categories, like grasping, punching and slapping, are initially learned as actions that bring about distinctive patterns of tactile feedback: the rich pattern of feedback produced by a grasp is different from that produced by a slap, and again from that produced by a punch. To train the simple action network, a simple rigid object (a cylinder, illustrated in Figure 5a) is presented to the reach network in a random location, generating a visual representation. The reach network computes a goal motor state from this visual representation, which is passed as input to the simple action network, which produces a perturbation of this goal state. This perturbation is annealed with noise, which is progressively reduced to zero during training. The feedback controller moves the hand towards the perturbed goal state; when it attains a certain distance from the target, the perturbation is removed, and the hand approaches the actual goal state.



(a) (b)

Figure 5: Objects created for the simulations. (a) A cylinder (for grasping, punching and slapping); (b) from left: a lever (for bending); a door (for opening); a compressable object (for squashing).

From time to time, the perturbation applied results in richer tactile reward signals than those used to train the reach network. Different rich tactile signals result from particular perturbations. Some perturbations result in a grasp or near-grasp, which generates one class of tactile stimulus. Others result in slapping movements, which generate another, different, class of tactile stimuli, or in punching movements, which generate another distinct class of tactile stimuli. (These rich stimuli are almost never generated through pure feedback control, because they result from special trajectories.) When a rich tactile stimulus is generated, copy and learning operations take place in the simple action network which are analogous to those in the reach network. First, the tactile stimulus is copied to an area holding 'motor programs'. Second, the simple action network is trained to map the current goal motor state, *plus the currently active motor program*, onto the perturbation which resulted in the reward. After this learning, activating a specific motor program will generate an action with a characteristic trajectory. We envisage motor programs competing with one another, with a single winner being selected.

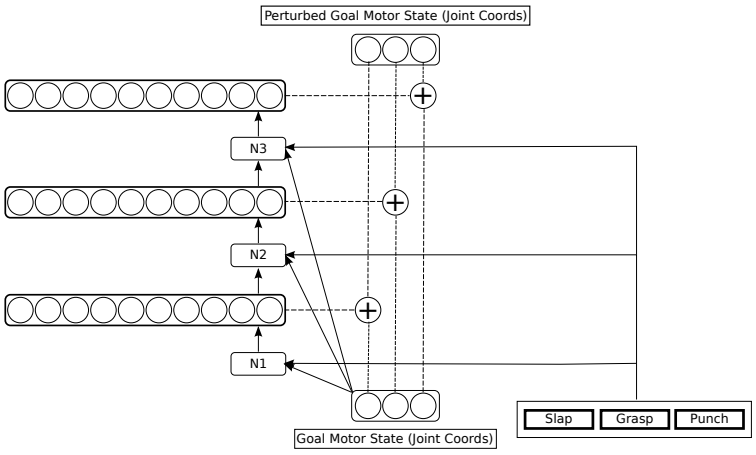



Figure 6: Detailed architecture of the simple action network

In the simple action network, the three motor components of a perturbation are computed one by one, in the three networks labelled N1, N2 and N3 in Figure 6. This is because there are typically several possible perturbations which result in any given tactile reward signal: the network needs to select one of these, and selection of the different components of a perturbation cannot be performed independently. So the network N1 computes the alternative possible values for the first component of the perturbation, then selects one of these, and passes the selected value to the network N2 as input, and N2 performs a similar operation. In each network, ten output neurons are used to encode this distributed coding with each representing a particular discrete value in the range of possible joint angles. The neuron with the highest activation is chosen to be the 'best' motor program and the activation of the nearby neurons is used to alter the value slightly. All of the neurons outside of this region are inhibited and are not considered when decoding this perturbation component.

Note that the simple action network must execute in parallel with the simple reach network. It modulates the behavior of the simple network, in a manner reminiscent of Brooks' (1991) subsumption architecture. In order to execute a simple motor program, it is important that the whole simple action circuit is enabled, or turned on. Accordingly, while different motor programs provide different input to the simple action network, they also uniformly generate a control signal to enable the network they provide input to. This control signal is shown in Figure 3 by the unfilled arrow leading from the simple motor program medium to the simple action network.

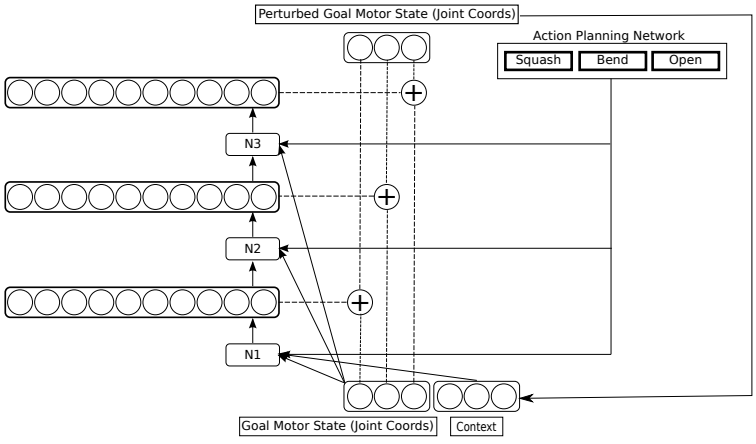



Figure 7: Detailed architecture of the causative action network

The final network to be trained is the **causative action network** (see the blue part of Figure 3, and Figure 7). This network implements our model of causative actions. Our key proposal is that above the

simple action network there is a higher-level network trained from still more sophisticated sensory signals, which derive not from the tactile system, but from a high-level perceptual module which can classify arbitrary actions taking place in the external world, relying mainly on vision and hearing rather than touch. There is a well-studied perceptual module of this kind in the brain, implemented in a pathway from sensory cortices (in particular visual cortex) through the superior temporal sulcus (STS) and inferior parietal cortex to the premotor cortex (see e.g. Keysers and Perrett, 2004). When an agent allocates attention to an external object, representations in this pathway encode the actions *of this object*, in increasingly complex ways. This action classification pathway is normally thought of as being engaged when an agent is passively observing the external world. But consider what happens when the agent is attending to an external object *as a target*, while directing the hand towards it along a particular trajectory. Any actions regularly evoked in the action recognition pathway in this scenario are likely to be actions *caused by the hand's movement*. We propose that during action execution, action signals evoked in the action recognition pathway are hardwired to function as reward signals, which train the causative action network to bring about particular distal actions in the world.

Training in this higher-level motor circuit involves presenting different articulated objects to the system, which can undergo various different changes in configuration. Objects are of three types: One is a lever that can pivot around a joint, and can be bent; one is a hinged door in a plane, which can be pushed open; one is a pair of horizontal plates connected by a spring, which can be 'squashed' by pushing down on the top plate. These objects are illustrated in Figure 5b.

Training the causative action network again proceeds by random generation of perturbations to the goal motor state delivered by the reach network. In this circuit, *sequences* of perturbations are applied, to generate still more complex trajectories. (This is depicted in Figure 7 by a recurrent input, though in our implementation we 'unroll' this recurrence and generate exactly two perturbations.) Some of these sequences cause particular patterns of movement in the target object, which are interpreted as external actions by the action recognition system. Activation of an action representation in the action recognition system when performing an action on a target object is hard-wired to generate a reward signal. This signal has two effects. First, the observed action is copied to a medium in which action plans are held (the action planning system). Second, the causative action network is trained to map the basic goal motor state delivered by the reach network onto the sequence of perturbations which led to reward. Note that the network also takes representations in the action planning system as input. After training, the causative action network can take a simple goal motor state, plus an action representation in the action planning system, and generate a

sequence of perturbations which will lead to observation of the planned action on the attended target.

This network enables a rich repertoire of actions to be learned. It preserves Hommel *et al.*'s idea that action representations are organised around their perceptual effects. But since the action recognition network generates rich, high-level perceptual signals, a correspondingly rich set of motor programs can be established. At the same time, the basic mechanisms through which learning happens are the same as in much simpler motor learning systems.

Part of the design of the causative action circuit is that 'cause' is a motor program in its own right, which competes within the motor program selection system against regular motor programs like 'grasp' and 'slap'. One important difference is that the 'cause' action enables the causative actions network rather than the simple action network, but other than that it counts as a regular motor program. This raises some important questions about how causative actions are planned and executed. When an agent decides to perform a causative action, presumably he has some particular caused action in mind. But at the time of planning, this caused action is in the future: minimally, the agent must bring his hand into contact with the target object before he can cause it to move in any way. In order to cause a particular action in a target object, the trajectory of the hand towards the object must often be biased from the very start: for instance, to cause an object to squash, the hand must approach the target from a particular direction, and with particular force. So the movements which bring about the caused action must be initiated some time before the action is perceived.

Our way of addressing this issue in the network is to activate the motor correlates of perceived actions in the medium holding *planned actions*, rather than in the medium of regular motor programs like 'grasp' and 'slap'. An underlying assumption in our model is that an agent brings about actions through planned sequences of sensory or motor operations (for details see Knott, 2012). We also assume that planned sequences are selected as wholes, and that the component actions in a planned action sequence are active in parallel in the working memory medium where actions are planned. (This assumption is well supported by single-cell recordings in monkeys; see e.g. Averbeck *et al.*, 2002.) When the causative actions network is exploring causative actions, it will activate the 'cause' motor program experimentally, and choose a random sequence of perturbations. In some cases, this results *some time later* in activation of an action in the action recognition system: say 'squash'. This observed action activates a corresponding planned action. Additionally the **plan activation network** (see the bottom right of Figure 3) learns that the sequence 'cause', 'squash' is a good one to execute on the category of object currently present, so that when a similar object is presented in future, it will tend to activate this planned sequence. Now consider what happens

when the planned sequence is executed. The agent first executes the motor program 'cause'. This enables the causative action network, which generates a sequence of perturbations. Crucially, the causative action network also takes input from the planning medium in which the caused action ('squash') is active as part of the planned sequence. So as soon as it is initiated, the network is configured to generate the perturbation sequence which led to the caused action, even before this action actually occurs.

The key mechanism enabling causative actions to be executed is one which activates a sensory representation (the squash action) *as a goal* some time before it is evoked as a sensory stimulus. Note that something very similar happens in the other networks; for instance in the reach network the actual motor state where the touch sensation occurs is activated as a goal motor state. In the simple network this activation is possible because visual perception provides information about reward-associated motor states. In the higher-level causative actions network, the advance notification of reward comes from the working memory system which stores prepared actions. But the effect is much the same.

## 4 Training and testing of the network

We trained the reach network and simple action network on a cylinder presented at a range of locations, and the causative action network on the three articulated objects presented at the same locations. Training and testing of the reach network is described in detail in Lee-Hand *et al.* (2012); training and testing of the other two networks is described in detail in an appendix to the current paper (Lee-Hand and Knott, 2013).

The trained simple action network was tested by presenting a cylinder at a number of selected locations, activating a simple motor program at random (grasp, slap or punch), and observing how often the tactile stimulus associated with this motor program was produced. Results from these tests are summarised in Figure 8a. The causative action network was tested by presenting one of the articulated objects at one of a number of locations, and observing how often the network generated a series of perturbations that led to the action perception system registering the action appropriate for the object. Results of these tests are presented in Figure 8b. In general, the system was quite successful in producing motor actions with the expected perceptual consequences. The motor program network produced actions resulting in the expected tactile stimuli for an average of 86.25% of seen target locations and an average of 81.2% of unseen locations; the causative action network produced actions resulting in the target undergoing the expected action in an average of 86.7% of seen locations and an average of 83.8% of unseen locations. Illustrations of representative successful actions of each type are shown in Figure 9.

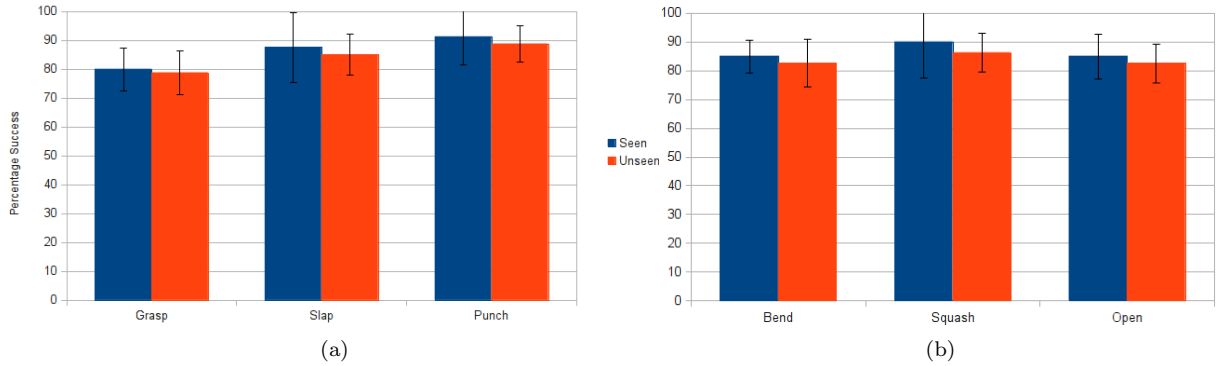



Figure 8: (a) Results from testing the simple action network. (b) Results from testing the causative action network. Error bars show standard deviation for seen and unseen locations across 10 trials of 8 object locations.

## 5 Discussion of the network model

While there are many computational models of learning to grasp in the field of robotics, there are comparatively few models of how hand/arm actions are learned in infants. By far the most relevant is the neural network model of Oztop *et al.* (2004). This model focusses on infant learning of grasp actions directed at rigid target objects. One point of similarity already mentioned is that this model considers touch sensations to be intrinsically rewarding. Our model adopts this idea of 'the joy of grasping', but extends it in various ways, by envisaging a progression of increasingly elaborate sensory stimuli functioning as rewards in a succession of increasingly high-level motor networks. Another point of similarity is in the way hand trajectories are defined. Oztop *et al.*'s trajectories are defined by 'via-points' through which the hand must pass on the way to the target: their network produces via-points in the way our network produces perturbations. Via-points are represented in a target-centred coordinate system, and so are perturbations, which are specified as deviations from a motor-centred representation of the target location. But there is an important difference: Oztop *et al.*'s model computes trajectories through via-point *in advance*, while perturbations create trajectories 'on-the-fly'. Our model avoids precomputation of trajectories, as there is no evidence that this is done by humans or monkeys (see again Cisek 2005). But there is a tradeoff: using perturbations to generate trajectories is much less precise than using via-points. How to produce precise trajectories without precomputing them is still an open question.

The most novel aspect of our model is its account of causative motor actions. The closest relevant work in this regard is the model of Arbib *et al.* (2009), which explores how an agent learns to interact with objects

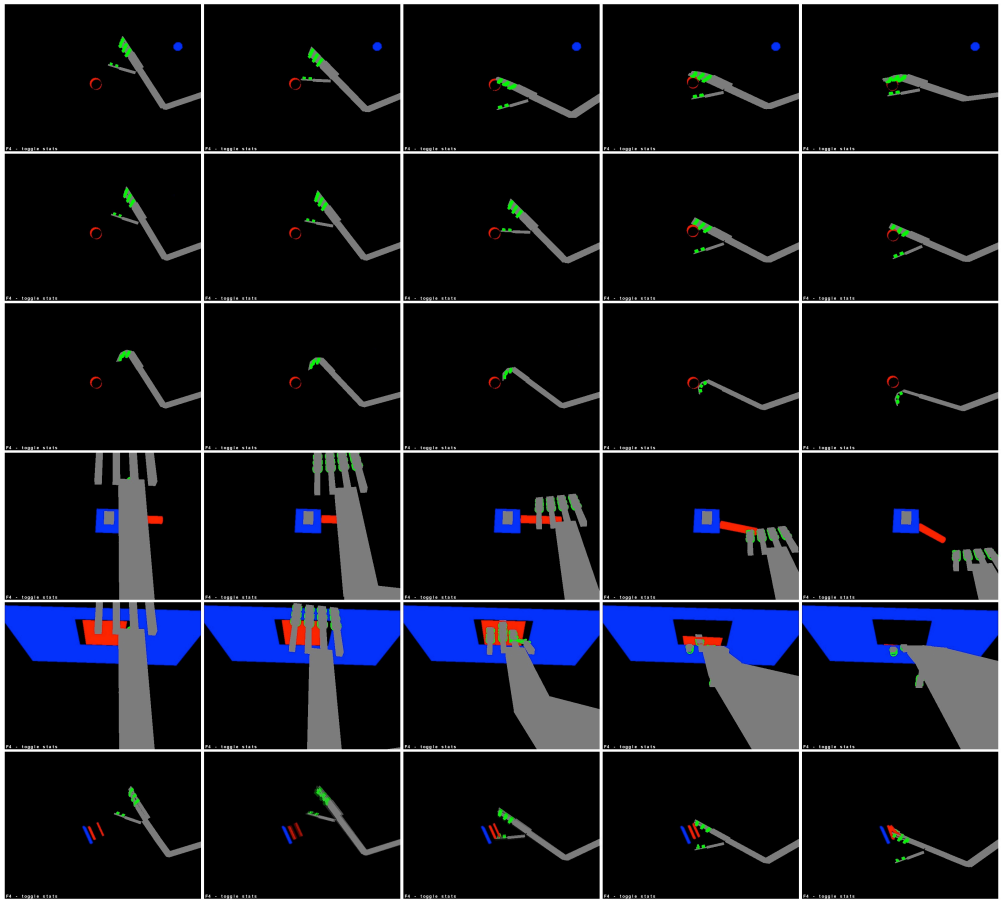


Figure 9: Learned actions. From top: grasping, slapping and punching a cylinder; bending a lever, opening a door and squashing a sprung plate. These sequences are taken from the latter stages of each action, when the hand makes contact with the target.

as tools. When an agent holds a tool, representations in the motor system often encode the actions *of the tool* rather than of the hand (see again Umiltà *et al.* 2008): this model aims to explain this phenomenon, which is closely related to the phenomenon we study in the current paper. The main difference in our model is that it does not only consider the actions brought about on an object when the hand is in contact with it; it also addresses how to configure the approach trajectory of the hand so as to achieve a particular action, and it can learn movements producing actions that occur after the hand loses contact with the target. Arbib *et al.*'s model focusses on how a tool becomes incorporated into an agent's motor representation of his hand as an effector, while ours focusses on how an effector can bring about movements in objects even without achieving a stable grasp on them. How to create a model that addresses both issues is again an open question.

Our model makes various novel predictions which could be tested experimentally. For one thing, it

predicts that the goal motor state of an effector can be perturbed while it is moving towards a target. We know something about the neural areas where motor goals are evoked (see again Cisek 2005); it would be interesting to examine whether changes in these representations during action execution influence the trajectory of the effector. Another interesting prediction is that the neural circuit controlling causative actions like bending and squashing is partly separate from, and more complex than, the circuit controlling simple actions like grasping and slapping. A specific prediction is that the former circuit involves action-recognition machinery in the superior temporal sulcus in a way that the latter does not. These predictions could readily be tested in imaging experiments.

## 6 Causative motor actions and causative sentence structures

In the remainder of the paper, we return to a discussion of causative structures in language. Recall that some transitive verbs can undergo the causative alternation (e.g. *John bent the lever*, *The lever bent*) while others cannot (e.g. *John grabbed the lever*, **The lever grabbed*). Syntacticians have accounted for the difference between these verb types by positing differences in the underlying structure of the transitive sentences in which they appear: the underlying structure of *John bent the lever* features an explicit causal verb (*John CAUSED [the lever bent]*), while that of *John grabbed the cup* does not. As already mentioned, this is an appealing syntactic model, but many questions remain about the nature of the hypothesised causative construction. In this section, we consider whether our computational model of causative actions can shed any light on these questions. In Section 6.1 we introduce a particular version of the syntacticians' proposal in some more detail; in Section 6.2 we introduce a general hypothesis about how underlying syntactic representations relate to sensorimotor processing, and in Section 6.3 we apply this hypothesis to the case of causative sentences.

### 6.1 A Minimalist model of the causative alternation

To express a syntactic model of the causative alternation in more detail, we will adopt Chomsky's Minimalism (1995) as a syntactic framework. In Minimalism, the underlying syntactic structure of a sentence is called its **Logical Form** (**LF**). LF is conceived as the level of syntactic representation that communicates with semantics: it includes information that is left implicit in the surface Phonetic Form (PF) of sentences. While languages vary a great deal in PF structure, the Minimalist assumption is that LF structure is relatively invariant over translation, capturing generalisations between languages that are not visible at PF. The principles that govern how LF structures are formed are assumed to be largely innate; what infants have to

learn is how to map LF structures onto the language-specific PF structures in their mother tongue.

The LF structure of a clause is recursively formed from elements called **X-bar schemas** (**XPs**), joined together in a largely right-branching way. The LF structure of the simple transitive sentence *I grabbed the cup* is shown (very schematically) in Figure 10a.[1] The X-bar schemas are indicated with boxes in the figure. We will use Figure 10a to briefly introduce some of the features of the Minimalist model.

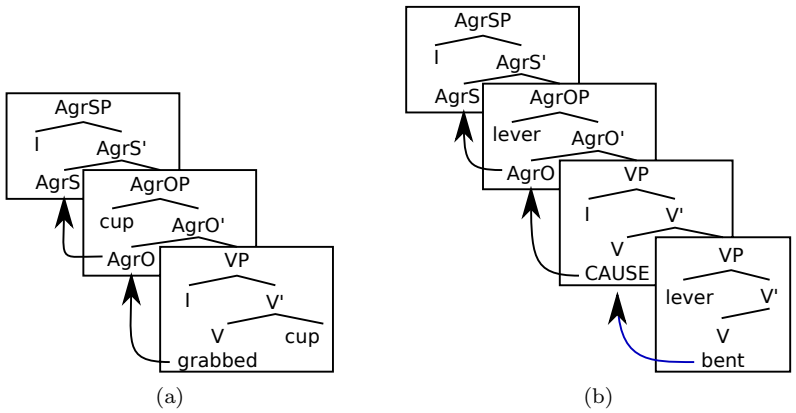



Figure 10: (a) LF of *I grabbed the cup*. (b) LF of *I bent the lever*. Arrows depict head-raising operations.

Note that the subject (*I*) and object (*cup*) each appear in two positions at LF structure. In the Minimalist model they originate at the lower positions within VP: at these positions the verb (*grabbed*) assigns the semantic role AGENT to *I* and PATIENT to *cup*. They then raise to their higher positions to be assigned 'Case', through a process we will not describe here. We will focus on what happens to the verb, which also appears in multiple positions at LF. This happens through a process called **head movement**, which is indicated by the arrows in Figure 10a. The verb originates at the head of VP, and moves to successively higher head positions in projections called AgrOP and AgrSP. In the Minimalist model, head movement is the device that explains how the verb can carry inflections agreeing with the subject and object. The 'low' verb position in VP only holds the meaning of the verb stem *grab*; the semantics of its agreement inflections are held at the heads of AgrSP and AgrOP, whose specifiers are the higher subject and object positions respectively. The verb has to move to these positions to be able to carry agreement inflections. Head movement is also the device that allows the inflected verb to appear at different surface positions in a sentence in different languages. It creates alternative possible surface positions for the verb: it can be pronounced 'high', as in VSO languages, or 'low', as in SOV languages. The multiple positions of the subject

[1] Our example sentences feature first-person subjects because our sensorimotor model experiences actions from the perspective of the agent; this will later allow us to match units in LF structure directly to operations in the sensorimotor model.

and object at LF similarly provide alternative surface positions for these constituents; in combination, the alternative positions of subject, verb and object allow accounts of all possible constituent orderings found in the world's languages (SVO, VSO, SOV and so on).

The LF structure of *I bent the lever* is shown in Figure 10b. The key difference here is that there are two VPs. The higher VP is headed by the verb *CAUSE*; the lower, introduced as the complement of the higher, is headed by the verb *bent*. The NP *lever* appears as the subject of this lower VP, in the structural position where it is established as the AGENT of *bent*. The key proposal in the syntactic model is that *bent* raises at LF to the position occupied by CAUSE, and at PF it can be pronounced at this higher position. In English, where the subject is pronounced high and the object low, this yields the surface ordering *John bent the lever*. This type of head-raising is a little different from the kind illustrated in Figure 10a: it involves raising a verb to combine with a higher verb, rather than raising a verb into an empty position. To indicate this, the body of the arrow is shown in blue. But the movement operation is still a variety of head-raising, so this movement-based account of the causative alternation is still relatively economical.

### 6.2 A general sensorimotor interpretation of Minimalist LF structures

As mentioned in Section 1.3, there is a growing body of evidence that representations of language recruit representations in the sensorimotor system (see e.g. Barsalou 2008; Glenberg and Gallese 2012). We will adopt one particular proposal about how syntactic representations connect to the sensorimotor system, which was introduced by Knott (2012). This proposal assumes a Minimalist model of syntax. Since Minimalism posits a level of syntactic structure (LF) that is relatively language-independent, it allows the formulation of particularly strong hypotheses about the relation between syntax and the sensorimotor system. Knott's proposal is that the LF structure of a sentence describing a concrete episode in the world can be interpreted *as a direct description of the sensorimotor process through which this episode was experienced.* This proposal offers a new perspective on Minimalist nativism: it sees the commonalities between languages expressed at LF as due not to innate *language-specific* knowledge, but to the fact that syntax directly references sensorimotor routines, which are the same for speakers of all languages.

In Knott's model, the recursive structure of LF mirrors a similarly recursive structure in sensorimotor routines. Knott assumes Ballard *et al.*'s (1997) account of sensorimotor processing, which posits that sensorimotor routines are organised into well-defined sequences of attentional or motor operations called 'deictic operations'. The key idea in Knott's proposal is that the LF structure of a concrete sentence describes a sequence of deictic operations—specifically, that each XP in the structure describes a single operation.

Knott (2012) presents a detailed model of episode perception, in which experiencing a cup-grabbing episode involves a well-defined sequence of sensorimotor operations. Each operation creates the context in which the next operation can be executed. We will illustrate the sequence with reference to the processing in our model of motor control when it executes the action of grasping an object. The initial operation is a *decision to act*, which corresponds to the action of engaging the network as a whole. This operation basically establishes the agent *as an agent*, rather than a passive observer, by routing incoming sensory inputs into circuits that generate motor plans and motor outputs. The next operation is activation of a visual representation of the target object. This would normally involve attentional selection of a single target from many candidates, but our model assumes this process has already happened: the initial representation in our network is of the shape, location and category of a single object. The final operation is the activation of the simple motor programme 'grasp', which results in the production of a motor movement. In Knott's model, this movement activates a second representation of the agent, this time as a temporally extended pattern of motor signals (not modelled in our implementation). At the end of the grasp action, when the hand makes contact with the target object, a second representation of the target is activated, this time in the haptic system: the rich tactile signal allows the hand's current position and shape to convey information about the position and shape of the target object.

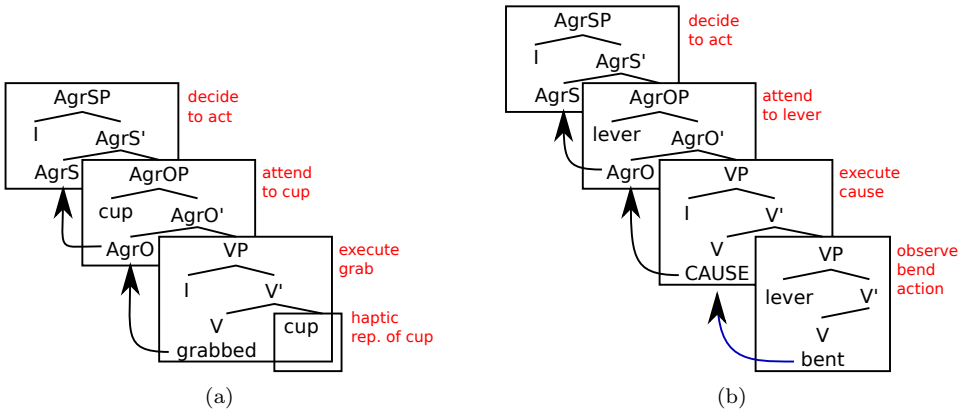



Figure 11: Sensorimotor interpretations of the LF structures in Figures 10a and 10b

The sequence of sensorimotor operations involved in experiencing a cup-grabbing action (as an agent) map well onto the LF of the sentence *I grabbed a cup*, as shown in Figure 11a. The sensorimotor interpretation of LF is shown in red. The AgrSP projection that hosts the subject *I* and the verb's subject agreement features describes the agent's decision to act, which establishes him as the agent of the action to be experienced. The

AgrOP projection that hosts the object *cup* and its agreement features describes the action of attention to the cup, which produces a representation of the cup as a goal motor state. VP describes the activation of the motor programme *grasp*, which evokes a second representation of the agent, this time as a pattern of movement. The complement of VP, *cup*, describes the haptic representation of the cup activated when the grasp routine is completed. Note that the two subject and object positions in the LF structure correspond to the two representations of the agent and patient, which are activated in different sensorimotor modalities at different points during execution of the grasp routine.

Most relevantly for the current paper, Knott also proposes a sensorimotor interpretation of the head-raising operation that moves the verb *grabbed* from the head of VP to the heads of AgrSP and AgrOP. He argues that LF structures describe sensorimotor routines as replayed from working memory, rather than as they are experienced in real time. In Knott's model, an agent represents an episode in working memory as the sensorimotor sequence through which it was experienced. This sequence is stored in the prefrontal medium that holds *planned* action sequences. This means that the routine through which an episode was experienced can be *replayed*; see Takac and Knott (2013) for a model of this process. To generate a sentence describing an experienced episode, the associated routine is replayed in a special mode in which sensorimotor signals can activate surface linguistic forms (see Takac *et al.* 2012 for a computational model of sentence generation making use of this idea). In the model of motor control presented in the current paper, the medium storing a planned sequence of sensorimotor operations is the action planning medium. As discussed in Section 3, this medium stores a sequence of actions by representing actions in parallel. When a sensorimotor routine is replayed, the planning medium holds tonically active representations of each action in the routine throughout the replay process. To explain head raising, Knott proposes that the head of an XP describes a sensorimotor operation *in the planning medium*, rather than as it occurs transiently during actual or replayed experience. The head of VP describes a *planned* motor action, which is signalled linguistically by a verb stem; the heads of AgrSP and AgrOP describe *planned* actions of attention to the agent and patient of the grab action, which are signalled linguistically as agreement inflections. (There is good evidence that production of verbs and inflections is associated with activity in dorsolateral prefrontal cortex, the same area that holds planned sensorimotor routines; see e.g. Perani *et al.* 1999; Shapiro and Caramazza 2003.) In Knott's account, the reason why the verb and its inflections can appear at any head position is that the sensorimotor signals from which they are read are active tonically, and in parallel, during the replay of the sensorimotor routine.

### 6.3 Applying the interpretation to causative sentences

In a syntactic model of the causative alternation, the LF of *I bent the lever* differs in two respects from that of *I grabbed the cup*. Firstly it features two VPs in right-branching sequence, the higher one headed by the verb *CAUSE*, and the lower by the verb *bend*. Secondly, the lower verb raises to the position of the higher verb through a special type of head-raising, moving a verb into a position already occupied by another verb. Knott's sensorimotor interpretation of LF includes a general proposal about the interpretation of a right-branching structure of XPs, and a general proposal about the mechanism responsible for head-raising. When applied to the proposed LF representation of *I bent the lever*, it makes strong predictions about the difference between regular motor actions and causative motor actions, which can be directly tested in our computational model.

We first consider the right-branching structure of the LF representation. Knott proposes that a right-branching structure of XPs at LF describes a replayed sequence of sensorimotor operations. His proposal thus predicts that executing a causative action involves two operations occurring in sequence: firstly the execution of a 'cause' action, and secondly, the experience of the caused action (in our case 'bend'). Does the execution of a causative action in our network model have this sequential character? It certainly does. Once the network is engaged, and a target is selected, the next thing that happens is that the CAUSE motor programme is executed. This motor programme generates a movement that some time later results in the target object undergoing an action, which is registered by the action perception system. The sequence of sensorimotor operations involved in executing the action of bending a lever maps closely onto the right-branching XP structure of *I bent the lever*, as shown in red in Figure 11b.

We next consider the sensorimotor interpretation of the head-raising operation moving *bent* onto *CAUSE* in the syntactic model. Knott's account of head-raising hinges on the proposal that heads signal sensorimotor operations in a planning medium, where they are active throughout a replayed sensorimotor routine, rather than just at specific points. This means that linguistic signals of sensorimotor operations can be read out at the head of any XP in the right-branching clause structure. This proposal straightforwardly provides an account of how the lower verb *bent* can appear at the position of the higher verb *CAUSE*. Executing the action of bending a lever involves a sequence of four sensorimotor operations: deciding to act, attending to the lever, activating the 'CAUSE' motor programme, and activating the 'bend' action. An experienced lever-bending episode is stored in working memory as a planned sequence of these four operations in the action planning medium: since the representations in this medium are active in parallel when this sequence is replayed, they can all be pronounced in any head position, so it is perfectly possible to pronounce *bent*

at the higher V head. But recall that we also want to account for the special nature of the operation that raises *bent* into the position occupied by *CAUSE*, which is different from the operation that raises a verb to a higher agreement head. Again there is something in the model of motor control that accounts very neatly for this. The key point is that the action planning medium is involved in a special way in the circuit that *executes* a causative action, as well as in the machinery that produces an output sentence. Recall that in order for a 'CAUSE' operation to be able to bring about a perceived action, the causative action network must receive input not only from a perceptual representation of the target object, but from a representation in the planning medium, of the action that is to be brought about. So even during execution of a causative action, the parallel activation of 'CAUSE' and 'bend' in the planning medium plays a crucial role in the generation of an appropriate motor action: these action representations are not only active in parallel, but *influence motor output in parallel* during action execution.

In brief, Knott's sensorimotor interpretation of LF extends well to causative actions. The causative LF structure motivated on purely syntactic grounds, when interpreted in sensorimotor terms according to Knott's proposal, identifies elements of the computational model of causative actions which have their own independent motivation as components of this model.

## 7 Summary and future work

We had two goals in this paper: first, to introduce a computational model of causative actions, and second to consider whether this model sheds any light on the syntactic model of causative sentences proposed by syntacticians to account for the causative alternation. The computational model of causative actions has several novel features in its own right. It is based on the novel proposal that observed external actions can function as rewards to train the motor system, and it makes a novel proposal about how a representation of the planned perceptual effect of a movement can influence the movement before it is actually perceived. It also suggests how Hommel *et al.*'s proposal that actions are defined as their perceived sensory consequences can be implemented for actions of different degrees of complexity. Finally, it implements a novel method for generating trajectories using perturbed goal motor states. But we also believe the computational model provides a useful framework within which to understand a syntactic model of the causative alternation. The model of the LF of causative sentences like *I bent the lever* includes an explicit action of causation, but there is nothing in the model to explain the complexity of the relation between causative movements and caused actions in real movements. We suggest that an embodied account of syntax, where LF structures are interpreted as direct descriptions of sensorimotor routines, fleshes out L&RH's syntactic proposal in a very

informative way: if we think of the verb *CAUSE* as denoting the 'cause' motor programme in our network model, we can interpret the syntactic relations linking *CAUSE* to its complement VP as direct references to the motor mechanisms that actually produce causative actions. Our computational model identifies both sequential and parallel aspects of these mechanisms. The structural relationship between the VPs headed by *CAUSE* and *bend* references the sequential relation between activation of the motor programme 'CAUSE' and activation of the perceived action 'bend'; the raising of *bend* to adjoin to the higher verb *CAUSE* references the fact that the action 'bend' *within the planning medium* influences motor output in parallel with the motor programme 'CAUSE'.

We conclude by discussing some predictions of our model of causative sentences. We propose that nested VPs in causative sentences encode causative motor routines, and therefore predict that difficulties processing nested VPs may stem from motor deficits. An interesting study bearing on this prediction was conducted by Roy *et al.* 2013. This study tested the motor abilities of children with specific language impairment (SLI), a condition whose core deficit is in the ability to process nested syntactic structures. The motor task was to execute a causative action, namely moving a bottle of known weight. Children with normal language abilities execute this action in a smooth motor movement whose kinematics is tailored from the outset to the required displacement of the bottle; but SLI children execute the action in two distinct movements, one to reach the bottle, the next to make it move. Roy *et al.* propose that the SLI children's deficit in generating nested syntactic structures may stem from a deficit in the system that generates 'nested motor actions'. Our motor control network provides a detailed model of exactly this system: indeed, it produces the same smooth movements that children with normal language produce. It would be intriguing to explore whether damage to the network model could reproduce the kind of movements and syntactic structures produced by SLI children—and conversely, to investigate whether SLI children have difficulty processing transitive sentences that undergo the causative alternation.

## References


M Araki. PID control. In H Unbehauen, editor, *Control Systems, Robotics and Automation Vol. II*. 2006.

M Arbib, J Bonaiuto, S Jacobs, and S Frey. Tool use and the distalization of the end-effector. *Psychological Research*, 73:441–462, 2009.

B Averbeck, M Chafee, D Crowe, and A Georgopoulos. Parallel processing of serial movements in prefrontal cortex. *PNAS*, 99(20):13172–13177, 2002.

D Ballard, M Hayhoe, P Pook, and R Rao. Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4):723–767, 1997.

L Barsalou. Grounded cognition. *Annual Review of Psychology*, 59:617–645, 2008.

R Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

N Chomsky. *The Minimalist Program*. MIT Press, Cambridge, MA, 1995.

P Cisek and J Kalaska. Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, 45:801–814, 2005.

D Dowty. *Word Meaning and Montague Grammar*. Reidel, Dordrecht, 1979.

A Glenberg and V Gallese. Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, 48:905–922, 2012.

B Hommel. Inverting the Simon effect by intention. *Psychological Research*, 55:270–279, 1993.

B Hommel, J Müsseler, G Aschersleben, and W Prinz. The theory of event coding (TEC): A framework for perception and action learning. *Behavioral and Brain Sciences*, 24:849–878, 2001.

M Kawato, K Furawaka, and R Suzuki. A hierarchical neural network model for the control and learning of voluntary movements. *Biological Cybernetics*, 56:1–17, 1987.

C Keysers and D Perrett. Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences*, 8(11):501–507, 2004.

A Knott. *Sensorimotor Cognition and Natural Language Syntax*. MIT Press, Cambridge, MA, 2012.

J Lee-Hand and A Knott. Training and testing of a neural network model of motor control. Technical Report OUCS-2013-11, Dept of Computer Science, University of Otago, Dunedin, 2013.

J Lee-Hand, T Neumegen, and A Knott. Representing reach-to-grasp trajectories using perturbed goal motor states. In *Proceedings of the Pacific Rim Conference on Artificial Intelligence (PRICAI)*, pages 250–261, 2012.

B Levin and M Rappaport Hovav. *Unaccusativity: At the syntax-lexical semantics interface*. MIT Press, Cambridge, MA, 1995.

K Matsumoto, W Suzuki, and K Tanaka. Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science*, 301:229–232, 2003.

G McKoon and J Love. Verbs in the lexicon: Why is hitting easier than breaking? *Language and Cognition*, 3(2):313–330, 2011.

J Moore and S Obhi. Intentional binding and the sense of agency: A review. *Consciousness and Cognition*, 21:546–561, 2012.

T Neumegen. A computational platform for simulating reach-to-grasp actions: modelling physics, touch receptors and motor control mechanisms. MSc thesis, Dept of Computer Science, University of Otago, 2013.

E Oztop, N Bradley, and M Arbib. Infant grasp learning: a computational model. *Experimental Brain Research*, 158:480–503, 2004.

T Parsons. *Events in the Semantics of English*. MIT Press, Cambridge, Mass., 1990.

D Perani, S Cappa, T Schnur, M Tettamanti, S Collina, M Rosa, and F Fazio. The neural correlates of verb and noun processing—A PET study. *Brain*, 122:2337–2344, 1999.

W Prinz. Perception and action planning. *European Journal of Cognitive Psychology*, 9(2):129–154, 1997.

A Roy, A Curie, T Nazir, Y Paulignan, V des Portes, P Fourneret, and V Deprez. Syntax at hand: common syntactic structures for actions and language. *PLoS One*, 8(8):e72677, 2013.

K Shapiro and A Caramazza. Grammatical processing of nouns and verbs in left frontal cortex? *Neuropsychologia*, 41(9):1189–1198, 2003.

J Simon. Reactions towards the source of stimulation. *Journal of Experimental Psychology*, 81:174–176, 1969.

M Takac and Knott. A neural network model of working memory for episodes. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, Berlin, 2013.

M Takac, L Benuskova, and A Knott. Mapping sensorimotor sequences to word sequences: A connectionist model of language acquisition and sentence generation. *Cognition*, 125:288–308, 2012.

M Umiltà, L Escola, I Intskirveli, F Grammont, M Rochat, F Caruana, A Jezzini, V Gallese, and G Rizzolatti. When pliers become fingers in the monkey motor system. *PNAS*, 105(6):2209–2213, 2008.

**Figure Captions:**

Figure 1: Derivation of *John bent the lever* by movement form an underlying syntactic structure.

Figure 2(a): The hand/arm model used in simulations.

Figure 2(b): Detail of a single finger pad.

Figure 3: Architecture of the model control network.

Figure 4: Detailed architecture of the reach network.

Figure 5: Objects created for the simulations.

Figure 5(a): A cylinder (for grasping, punching and slapping);

Figure 5(b): from left: a lever (for bending); a door (for opening); a compressable object (for squashing).

Figure 6: Detailed architecture of the simple action network.

Figure 7: Detailed architecture of the causative action network.

Figure 8(a): Results from testing the simple action network.

Figure 8(b): Results from testing the causative action network. Error bars show standard deviation for seen and unseen locations across 10 trials of 8 object locations.

Figure 9: Learned actions. From top: grasping, slapping and punching a cylinder; bending a lever, opening a door and squashing a sprung plate. These sequences are taken from the latter stages of each action, when the hand makes contact with the target.

Figure 10(a): LF of *I grabbed the cup*.

Figure 10(b): LF of *I bent the lever*. Arrows depict head-raising operations.

Figure 11: Sensorimotor interpretations of the LF structures in Figures 10a and 10b.