Insights into event representation from a sensorimotor model of event perception

Alistair Knott, Martin Takac, Mark Sagar

Insights into event representation from a sensorimotor model of event perception

Alistair Knott
Soul Machines, Ltd
University of Otago, New Zealand
alistair.knott@soulmachines.com

Martin Takac
Soul Machines, Ltd
Comenius University, Slovakia
martin.takac@soulmachines.com

Mark Sagar
Soul Machines, Ltd
University of Auckland, New Zealand
mark.sagar@soulmachines.com

***Abstract*—In this paper we argue that a sensorimotor model of how an agent experiences events (both as an observer and as a participant) sheds useful light on the question of how to represent the syntax and semantics of sentences that report events. Our focus is on how to model the similarities and differences between sentences reporting change-of-state events and those reporting transitive and intransitive actions.**

***Index Terms*—Event representations, thematic roles, unaccusative verbs, event perception, visual object tracking, causation**

I. Introduction

Linguists have long been preoccupied with the question of how to represent the semantics of sentences that report events. It is well accepted that there are many qualitatively different *types* of event representation. For instance, some events involve volitional actions, while others are nonvolitional processes. Within volitional action events, there are intransitive actions (*Sally shrugged*) and transitive actions (*Jim grabbed a glass*). Within nonvolitional events, there are events where objects move (*The cup fell to the ground*) and events where objects change their intrinsic properties (*The glass broke*; *The door opened*). Some events can involve causative processes: these can be volitional (*Sally broke the glass*, which means 'Sally caused the glass to break') or nonvolitional (*The fire broke the glass*, which means 'the fire caused the glass to break'). Alongside this typology, sentences that report events can make use of a variety of syntactic structures. For instance, we can use active sentences (*Jim grabbed the glass*; *Sally broke the glass*), or passive sentences (*The glass was grabbed*; *The glass was broken*). The way semantic information is encoded in syntax also varies dramatically across languages, as well as within languages. A striking example of this involves the 'syntactic Case' assigned to noun phrases. In some languages, Case roughly distinguishes agent-like participants from patient-like participants of events. For instance, in an English active sentence reporting a volitional action, the agent receives **nominative** Case, whether the action is transitive or intransitive (*She chased Mary*; *She shrugged*), while the patient (if there is one) receives **accusative** Case (*Sally chased her*). In other languages, Case makes a different distinction, assigning the **ergative** Case to the agent of an intransitive sentence and also to the *patient* of a transitive sentence, and assigning **absolutive** Case to the agent of a transitive sentence. In Tongan, for instance, we have *Mele*$_{ERG}$ *danced*, and *Mele*$_{ABS}$ *hit Sione*$_{ERG}$ (*Na'e tau'olunga a Mele*; *Na'e taa'i e Mele a Sione*).

There are many ways to approach modelling such language patterns. In this paper, we argue it's helpful to consider the cognitive mechanisms through which events are *experienced* when devising an account of these patterns. All the events we have just mentioned are concrete: they are the kind of event that an observing agent can directly *perceive* taking place in her surroundings, through vision or other senses. (The observer can also experience volitional events in the motor modality, if she herself is the agent of the event.) The hypothesis we will explore is that the sensory and motor mechanisms through which events are experienced strongly determine how these events are *cognitively represented*—and through these represesentations, how they are reported in language. This hypothesis is an example of an 'embodied' model of language, of the kind that are the focus for the current workshop.

We'll begin in Section II by introducing the platform we have developed for building embodied models of language: the BabyX system. In Section III we will outline our basic approach towards modelling the experience and representation of events. In Sections IV and V we describe a particular model of event representations, which has the potential to model a range of event types, and a range of alternative syntactic encodings of event participants. In Section VI we describe a sensorimotor (SM) processing mechanism that can deliver these event representations.

II. BabyX: a platform for building an embodied model of language

To investigate embodied models of language and cognition, we have developed a sophisticated model of a human infant, called BabyX (Figure 1; see e.g. Sagar *et al.* [1]). The model was initially developed as an academic project; we are continuing to develop it as a R&D theme in a commercial company, Soul Machines.

The BabyX system is a blend of computer graphics/animation and neural network modelling. BabyX has a simulated body, implemented as a large set of computer graphics models, and a simulated brain, implemented as a large system of interconnected neural networks. She has simulated visual system, taking input from a camera pointed at the user,

Fig. 1. BabyX, interacting with one of the authors

and from the screen of a web browser page she and the user can jointly interact with. She also has a simulated motor system. This controls her head and eyes, so her gaze can be directed to different regions within her visual feeds; and it controls her hands and arms, so she can click and drag objects in the browser window (which is presented as a touchscreen in her peripersonal space). She can also perceive events in which the user moves objects in the browser window, as well as events where these objects move under their own steam.

A key goal in the BabyX project is to model the brain mechanisms that allow the baby to talk about the events she experiences—both those she perceives, and those she participates in as an agent. These models will be the focus for the current paper.

III. A GENERAL MODEL OF EVENT PERCEPTION AND EVENT REPRESENTATION

Experiencing an event takes time, whether it is being passively observed, or actively produced. A key assumption in our model is that the baby must produce a representation of an event *incrementally*, one component at a time, rather than all at once. (There are certainly fast ways of identifying the 'gist' of an event at a single moment, as shown by Hafri *et al.* [2], but these perceptual mechanisms do not deliver the rich, accurate event representations that are needed for a linguistic interface.) Our basic assumption is that the process of event perception is structured as a *sequence* of relatively discrete sensory and motor operations: and that at each step, the baby adds something to a **working memory** (**WM**) representation of the event that's under way. In this model, a WM representation of the event being experienced is authored progressively, as experience proceeds (see Takac and Knott [3]). When the process of experiencing the event is finished—which is normally when the event itself finishes—the WM representation of the event will be complete, and the complete event representation can be stored in longer-term memory (see Takac and Knott [4] for details of this process).

Our specific model of event perception is that it is structured as a **deictic routine**. The concept of deictic routines is due to Ballard *et al.* [5]. These researchers began by proposing that **deictic representations** play an important role in sensorimotor processing. A deictic representation in an agent's brain is a representation that is 'implicitly referred' to the momentary disposition of the agent's body towards the world. Most of the visual representations computed in the brain are 'deictic' in this sense, because they implicitly represent *the thing that the agent is currently looking at*: their content makes implicit reference to the agent's direction of gaze. (Representations in the object classification pathway are mostly referred even more specifically, to the point in the world which the agent is *fixating*.) Ballard *et al.* then define a **deictic operation**, which is an operation which *redeploys* the body's sensorimotor apparatus, to change or update these implicit references. The prototypical deictic operation is a saccade, that shifts the agent's gaze to a new location. Note that deictic operations *update* deictic representations, which always refer to the *current* disposition of the sensorimotor apparatus to the world.

Ballard *et al.*'s crucial insight is that deictic operations updating deictic representations are often determined by *current* deictic representations. This means that sensorimotor experience tends to comprise discrete *sequences* of deictic operations: that is, **deictic routines**. Our basic proposal is that the process of experiencing any concrete event takes the form of a deictic routine. In our general model, each deictic operation in an event-perceiving deictic routine is registered by adding material to the WM medium that holds event representations. Thus, as the routine progresses, an event representation is progressively built in this WM medium.

We have already specified this model in detail for events representing transitive actions, whether these are performed by the agent, or by some external agent being observed. We also have a detailed model of how the WM representation formed during event perception is reported in language. As Ballard *et al.* appreciated, deictic routines are potentially very useful in an account of how SM experience interfaces with language, because they essentially *discretise* relevant pieces of SM experience. However, Ballard *et al.* did not advance any particular model of the interface with language. Our aim is to advance a specific model of this interface.

Our proposal begins with the idea that WM representations are *prepared*, *replayable* deictic routines. Thus, when an agent creates a WM representation of an event, this representation allows her to *replay* the associated event-experiencing process—either by *producing* an event (if she is its agent), or by 'simulating' the experience process. We then make a proposal about the interface between WM representations and language: we propose that to produce a sentence that reports an event stored in WM, the agent *replays* the associated deictic routine (in simulation), in a special cognitive mode where the representations activated during replay can trigger output phonology. Details of these models can be found in Knott [6], with an implementation in Takac *et al.* [7].

The models we have produced so far only cover transitive volitional actions. In the current paper, we describe how we are extending these models, so they account for a range of event types other than transitive volitional actions. We retain the general proposal that events are experienced through deictic routines, which progressively populate a WM event representation. There are two things to add. One is a more elaborate

model of the WM medium that holds event representations. We will do this in Sections IV and V. The other is a more elaborate model of event-experiencing deictic routines, that highlights the points where the observer must decide what *type* of event is under way. We will do this in Section VI.

IV. An initial model of WM event representations

The WM event representations in our model are composed of various distinct fields. In our original model (Knott [6], Takac *et al.* [7]), there was one field for the agent of a transitive action event, and one for the patient; then there was a field holding a representation of the action itself, as shown in Figure 2(a). We proposed that the deictic routine through which a transitive event is experienced always begins with an operation 'attending to the agent'. If the agent is the observer herself, this operation involves activating the motor system; if the agent is some external actor, this operation involves activating an event-perception system. In either case, the operation produces a 'deictic' representation of the agent, in the medium shown in red in Figure 2(a): this representation is copied to the 'agent' field of the WM event. We also initiate a visual tracker on the agent (if it's an external object).

The observer is now in a position to attend to the patient. If she is performing the action herself, this involves attending to an object in her own peripersonal space. If she is perceiving an external actor, this involves attending to the object this actor is attending to, and/or reaching for: these processes involve a joint attention mechanism, and a mechanism that extrapolates the trajectory of the observed actor's hand. In either case, she produces a deictic representation of the patient, which must this time be copied to the 'patient' field of the WM event. She also initiates another visual tracker on the currently attended object. The tracked agent and patient now supply bindings for the parameters of a dynamic model of transitive action perception, or of transitive action execution. In either case, the model generates a specific transitive action category, in a perceptual or motor medium (again shown in red in Figure 2(a)). In sum, the complete process of experiencing a transitive action event involves three deictic operations: attend-to-agent, attend-to-patient, and activate-transitive-action-category. Each of these operations fills in one of the fields of the WM event medium.

When all three fields are filled, and the event is complete, we can encode it in long-term memory. This involves creating a representation in the **LTM event encoding** medium (shown in brown in Figure 2(a)). This is an associative medium that learns representations of commonly experienced event types, or significant token events. A key idea in our model is that this LTM medium uses 'place-coded' representations of semantic event roles: for instance, 'John-as-agent' occupies a different medium from 'John-as-patient'; see Takac and Knott [4] for details. Importantly, the LTM event encoding medium supports *queries*: for instance, if we train it to encode the event *Mary chased John*, we can query it with partial event representations like *Mary chased [X]* and retrieve the missing field.

V. An extended model of WM event representations

To extend our model to a wider range of event types, we need to modify the WM representation just introduced. The WM model shown in Figure 2(a) fudges some important linguistic issues. For one thing, it assumes that agents are always attended to first when recognising transitive actions. There is indeed good evidence that when watching transitive actions, observers reliably attend to the agent, and then the patient (see e.g. Webb *et al.* [8]). But passive constructions in language strongly suggest it is possible to recognise transitive actions without attending to the agent at all: I can recognise that *My bag was snatched* without noticing which agent did the snatching.

For another thing, the fields in Figure 2(a) aren't of much use in an account of how the semantic participants in an event are realised syntactically. The WM representation specifies two *semantic roles*, or in linguistic terms, **thematic roles**: but these don't map in a straightforward way onto syntactic positions. For instance, in an active sentence, the subject position reports the AGENT of the event, and the object reports the PATIENT, but in a passive sentence, the subject position reports the PATIENT. There is similarly no way to read out nominative and accusative Case from the WM representation in Figure 2(a): nominative Case is assigned to the AGENT in an active sentence, but to the PATIENT in a passive sentence.

As a final point, there is nothing in the Figure 2(a) representation to support change-of-state events, or causative events. And there is nothing at all to support an account of how Case is assigned in languages like Tongan, with ergative-absolutive Case marking. In sum, we need a richer WM event representation.

The extended WM event representation we propose is shown in Figure 2(b). There are two new media for holding event participants, which focus on different semantic properties, and which both provide input to the LTM event-encoding medium. We will introduce these in turn.

The causation/change area

The **causation/change area**, shown in blue, focusses on representing events in which objects change (as reported in sentences like *The glass broke* and *The spoon bent*), and causative processes that bring these changes about (as reported in sentences like *John broke the glass*, or *The fire bent the spoon*). This area contains two fields, which are each defined as a cluster of related concepts.

The **changer/attendee** field represents an object that undergoes a change, either in location (for instance an object that moves), or in intrinsic properties (for instance an object that bends or breaks). This field can also be used to represent the agent of an intransitive volitional action, such as a shrug or a smile. Such actions bring about changes to the configuration of the agent's body: in this sense, the agent 'undergoes a change', just like a spoon that bends. (Note that *bend* can be a volitional intransitive action, as in *John bent down*.)

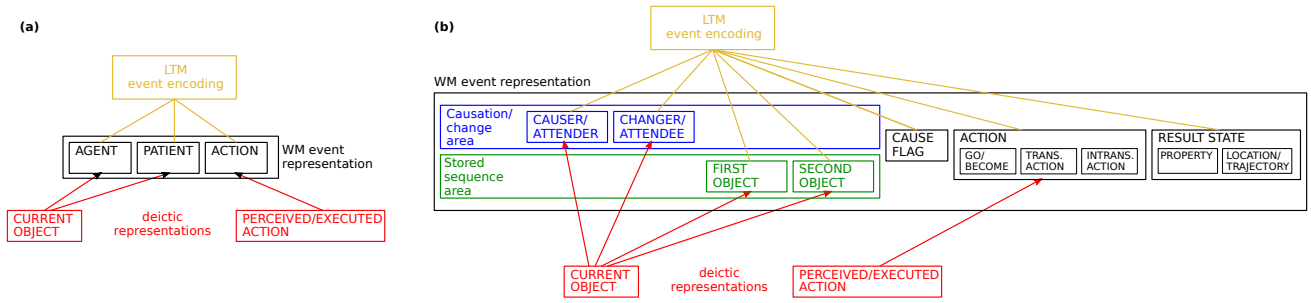

Fig. 2. (a) Our initial WM event medium, for representing transitive actions (black), and the circuits which populate the medium from deictic representations (red). (b) Our new proposed WM event medium.

The changer/attendee field also represents the patient of a transitive action. This patient isn't always changed: for instance, I can *touch* a cup without affecting it. But transitive actions *typically* change the target: so the roles of 'patient' and 'change-undergoer' often coincide. Our disjunctive definition of the changer/attendee field captures this regularity.

The **causer/attender** field represents an object that brings about a change in the changer/attendee. For instance, in *John bent the spoon*, it represents John, and in *The fire bent the spoon*, it represents the fire. By a similar disjunctive definition, this field also represents the agent of a transitive action: transitive actions needn't bring about changes on the target object, but they *often* do, so the agent is often a causer too.

In a sense, the causer/attender and changer/attendee extend the simple definitions of 'agent' and 'patient' in our original model in Figure 2(a). But there is one important difference: in the new scheme, the causer/attender field *doesn't have to be filled*. In our earlier model, the agent always coincided with the participant attended to first by the observer. In our current model, we capture this information separately, in the 'stored sequence' area (as we will discuss below). Allowing the causer/attender field to be blank lets us represent pure change-of-state events like *The glass broke*, which have no reference to a causer. They also let us represent passive events, like *John was kissed*, which have no reference to an agent.

The causation/change area makes useful generalisations over change-of-state events. Consider an event where a glass breaks, and another where some agency (John or the fire) causes the glass to break. We would like our LTM event-encoding medium to represent similarities between these: in particular, we would like its representation of the change that occurs to be the same. The causation/change area achieves this: if we store an event in which John breaks the glass, and then we query the LTM medium with the question 'Did the glass break?', we will get the right (affirmative) answer.

The causation/change area also provides a basis for an account of ergative and absolutive Case. As just outlined, the changer/attendee field holds the agent of intransitive event sentences, and also the patient of transitive event sentences, while the causer/attender field holds the agent of transitive sentences. If an event participant features as changer/attendee, it is therefore eligible for ergative Case, and if it features as causer/attender, it is eligible for absolutive Case.

The new WM event scheme shown in Figure 2(b) also includes some additional fields for representing change-of-state events. The 'action' field now includes a category of action called **go/become**. If the observer registers a change-of-state event, this category of action is indicated. (Note that the verb *go* can indicate a change in intrinsic properties (*John went red*) as well as a change in location (*John went to the park*.) We also include a new field called **result state**, that holds the state that is reached during a change-of-state event. This field has sub-fields for specifying object properties (such as 'red') and locations/trajectories (such as 'to the park').

Finally, the new WM scheme features a flag that indicates for change-of-state events whether a causal process bringing about the change-of-state is identified. This flag is set in events like *John bent the spoon* or *The fire bent the spoon*, but not in *The spoon bent*. Importantly, we assume that a causal process can be identified even if the causer object is not attended to. This gives us scope for representing passive causatives, such as *The spoon was bent*, which conveys that 'something caused the spoon to bend', without identifying that thing.

In relation to existing linguistic accounts of event structure, the causation/change area together with the 'cause', 'go/become' and 'result state' fields coincides quite closely to the well-known account of **unaccusative verbs** proposed by Levin and Rappaport Hovav (L&RH) [9]. An unaccusative verb is one that describes a change-of-state, roughly speaking: *bend*, *break*, *open* are prototypical examples. Unaccusative verbs often undergo the 'causative alternation', allowing sentences like *X bent Y* (meaning 'X caused Y to bend'). L&RH propose an underlying semantic structure for all unaccusative verbs: to illustrate, the semantics of the verb *break* is glossed as asserting a causal relation between two events:

[[X DO-SOMETHING] CAUSE [Y BECOME BROKEN]]

The fields in our WM event medium allow for exactly this analysis of unaccusative verbs.

The stored sequence area

The **stored sequence area**, shown in green, holds event participants in the order they were attended to. As noted, we now keep this information separate from encodings of

causality and change. There are two fields here, called **first-object** and **second-object**, which straightforwardly take copies of the first and second objects attended to. Note there is no second object in passives (*Mary was kissed*, *The spoon was bent*) and in pure change-of-state sentences (*The spoon bent*).

The objects occupying the 'first-object' and 'second-object' fields are semantically heterogeneous, just like those occupying the 'causer/attender' and 'changer/attendee' fields. But again, useful generalisations are captured across these categories. In particular, volitional agents of actions always occupy the first-object field, whether the action is transitive or intransitive, and whether it is causitive or not. We would like our LTM event-encoding medium to encode the volitional agent of actions in the same way, so we can query 'What did John do?', and retrieve all events, whether transitive or intransitive, causative or non-causative. Fields in the causation/change area can't provide this functionality, but the first-object field can do so.

Note also that the 'first-object' and 'second-object' fields provides a good basis for an account of nominative and accusative Case. Recall from Section I that the agent of active transitive and intransitive sentences receives nominative Case, as does the patient of passive sentences: the patient of active transitive sentences is the exception, in receiving accusative Case. In our model, if an event participant features as first-object, it is eligible for nominative Case, and if it features as second-object, it is eligible for accusative Case. These features also identify the (surface) subject and object of sentences: the participants receiving nominative and accusative Case appear as the subject and object of the sentence respectively.

The distinction between first-object and second-object also corresponds to a well-known classification of event participant roles—namely, that proposed by Dowty [10]. Dowty's interest is precisely in stating a general proposal about how semantic features of event participants determine the syntactic positions they hold within sentences (subject and object). It's not possible to formulate a precise rule about this—but Dowty suggests that 'cluster concepts' can be helpful in formulating the appropriate rule. He defines two cluster concepts: 'proto-agent' and 'proto-patient'. The proto-agent is defined via a cluster of agent-like features, including things like animacy, volitionality, sentience and causal influence. The proto-patient is defined via a cluster of patient-like features, including relative lack of movement, and the undergoing of state changes. Crucially, the participant that becomes the subject is the one that has the most agent-like features: for Dowty, participants are essentially in *competition* to occupy the subject position. In our model, this competition is an *attentional* competition: the participant attended to first occupies the 'first-object' field, and through this is selected as the grammatical subject.

VI. A model of event processing mechanisms

In this section, we outline a processing mechanism that can *construct* the kinds of WM event representation just introduced. As noted in Section III, this mechanism is structured as a deictic routine. The key novelty is that there are *choice points* at various places in this routine, where different types of event can be selected. Choices made are reflected in the fields of the WM event medium, and determine the course of the subsequent routine. Choices are made based on the outputs of visual mechanisms running in real time. We will first introduce these mechanisms, and then outline the deictic routine itself.

Visual mechanisms informing event perception

Central to our perceptual infrastructure are *two independent visual trackers*, configured to operate on different semantic targets. The **causer tracker** is set up to track the causer/attender; the **changer tracker** is set up to track the changer/attendee. A number of different classifiers then operate on the visual regions returned by these trackers (which we'll refer to as the **causer region** and **changer region** respectively).

Three mechanisms operate on the 'changer region' returned by the changer tracker. One mechanism is a regular **object classifier/recogniser**, which delivers information about the type and token identity of the tracked object (and also about its salient properties) to the 'current object' medium. A second mechanism is a **change detector**, comprising a **movement detector** (identifying change in physical location) and a **property change detector** (identifying change in the properties identified by the object classifier, including changes in body position). A third mechanism is a **change classifier**, that monitors the *dynamics* of the changer object in physical space and property space. If the changer object is animate, some dynamic patterns are identified by an **intransitive action classifier**, as changes that can be initiated voluntarily, like shrugs and smiles.

Two separate mechanisms operate on the 'causer region' returned by the causer tracker. One is an **animate agent classifier**, that attempts to locate a head and motor effectors (e.g. arms/hands) within the tracked region. If these are found, a **head tracker** and **effector tracker** are assigned to these sub-regions. If these sub-regions are found, a **directed attention classifier** operates on them, to identify salient objects near the tracked agent, based on the agent's gaze and/or extrapolated effector trajectories. A second mechanism is a **causative agency classifier**. This classifier assembles evidence that the tracked object is *influencing its surroundings*, either volitionally, through animate actions, or nonvolitionally, through perceived properties like heat or motion, or through learned knowledge about the object. The causative agency classifier monitors the dynamics of the tracked causer object, just as the change classifier monitors the dynamics of the changer object.

A final set of mechanisms operate *jointly* on the causer and changer regions returned by the two trackers. The first of these is a **transitive action classifier**. This classifies patterns of agent-like movement in the causer region, and patterns of pose in this agentlike object's hands, if these are being tracked. It also monitors movements of this agent's effectors *towards the changer region*, which is understood to be place attended to by this agent. This classifier can recognise directed actions like grabbing, slapping and punching, which are characterised

by particular patterns of hand pose and biological motion, and particular hand trajectories onto the assumed target. The second mechanism is a **causative process detector**. This system attempts to *couple* the dynamics of the causer object (delivered by the causative agency classifier) with the dynamics of the changer object (delivered by the change classifier). While these classifiers are configured to operate on the causer and changer objects *together*, we assume that after training, they can also operate on the changer object *by itself*. By this assumption, we can recognise a transitive action done on the changer object, or a causal process influencing the changer object, by attending to the changer object alone. This assumption is important in an account of passive transitive and causative sentences.

A deictic routine for event perception

Step 1 in our deictic routine is to attend to the most salient object in the scene, and to assign *both trackers* to this object. Assigning the changer tracker allows the object classifier to generate a 'current object' representation (the red box in Figure 2(b)).

At this point we begin deciding what kind of event the attended object is participating in. Our first decision is whether to copy the object representation to the the causer/attender field, or to the changer/attendee field. Evidence for the changer/attendee field is assembled by the change detector, which is referred to the attended object by the changer tracker. Evidence for the causer/attender field is assembled jointly by the directed attention and causative agency classifiers, which are both referred to the attended object by the causer tracker. If the object is established as causer/attender, we implement **Step 2a**; if it's established as changer/attendee, we implement **Step 2b**. In either case, the object representation is also copied to the 'first-object' field of the WM event.

In **step 2a**, we retain the causer tracker on the current object, and attempt to reassign the changer tracker to a *new* object. To do this, we consult the directed attention and causative agency classifiers, to seek objects that are the focus of joint attention, or directed movement, or causative influence. If we find a plausible candidate object, we attend to this object, and reassign the changer tracker to this object. The object classifier then produces a representation of this new object in the 'current object' medium, which is copied to the changer/attendee field of the WM event, and to the 'second-object' field. We can now deploy the two classifiers that operate jointly on the causer and changer regions: the transitive action classifier (which looks for actions done by the causer on the changer, such as 'Mary slapped the ball'), and the causative process detector (which looks for causative influences of the causer on the changer, such as 'Mary moved the ball down'). (Note that these classifiers can both fire, if the causative process also happens to be a transitive action, as in 'Mary slapped the ball down'.) If a causative process is identified, we set the 'cause' flag in the WM event, and also the 'go/become' flag (because what is being caused is a change). If not, we don't. If a change is being caused, we monitor the change to completion, and in a final step, we write the 'result state' it reaches to the WM event.

In **step 2b**, we have a changer object, but no causer. We stop the causer tracker, but maintain the changer tracker on the currently attended object. We are now set to execute three separate dynamic routines. One is the same change-detection routine that operates in Step 2a. Again, if a change is detected, we set the 'go/become' flag. In this scenario, we produce unaccusative sentences like 'the glass broke'. The other two routines are the transitive action classifier and causative process detector, configured to operate *just on the changer object*, to give passives, as described before. The causative process detector only runs if change is also detected, giving sentences like 'the glass was broken'. And the transitive classifier only runs if *neither* change or causation are detected (e.g. in 'the cup was grabbed') or if *both* are detected (e.g. in 'the cup was punched flat').

VII. Summary

In this paper, we have outlined the model we are developing for perceiving and representing events. Regarding representation, our model draws on L&RH's account of unaccusatives and causatives, and on Dowty's account of proto-agents and proto-patients. Our main novel proposal is that event participants should be *doubly represented* in event structures, using both of these schemes. We argue this has benefits not only for modelling the interface between WM event representations and language, but also for modelling the LTM storage of events, in a format that supports meaningful query operations. (In the appendix of the paper—an optional extra—we illustrate the coverage of our event encoding scheme, by showing how it represents the semantics of a range of sentence types.) Regarding processing, we propose a deictic routine incorporating a cascade of choice points, which allows events of all supported types to be progressively identified during SM experience. This routine was just introduced briefly here, but we hope to have conveyed its hierarchically branching temporal structure.

References

[1] M. Sagar, M. Seymour, and A. Henderson, "Creating connection with autonomous facial animation," *Communications of the ACM*, vol. 59, no. 12, pp. 82–91, 2016.

[2] A. Hafri, J. Trueswell, and B. Strickland, "Extraction of event roles from visual scenes is rapid, automatic, and interacts with higher-level visual processing," *Cognition*, vol. 175, pp. 36–52, 2018.

[3] M. Takac and A. Knott, "Working memory encoding of events and their participants," in *CogSci*, pp. 2345–2350, 2016.

[4] M. Takac and A. Knott, "Mechanisms for storing and accessing event representations in episodic memory, and their expression in language: a neural network model," in *CogSci*, pp. 532–537, 2016.

[5] D. Ballard, M. Hayhoe, P. Pook, and R. Rao, "Deictic codes for the embodiment of cognition," *Behavioral and Brain Sciences*, vol. 20, no. 4, pp. 723–767, 1997.

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

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

[8] A. Webb, A. Knott, and M. MacAskill, "Eye movements during transitive action observation have sequential structure," *Acta Psychologica*, vol. 133, pp. 51–56, 2010.

Sentence	causation/change		prepared-sequence		cause	go/become	action	result-state
	causer/ attender	changer/ attendee	first- object	second- object				
Mary jumped	Mary	-	Mary	-	0	0	jump	-
Mary kissed John	Mary	John	Mary	John	0	0	kiss	-
John was kissed	-	John	John	-	0	0	kiss	-
Mary went over the pond	Mary	-	Mary	-	0	1	-	over-the-pond
Mary jumped over the pond	Mary	-	Mary	-	0	1	jump	over-the-pond
Mary threw John over the pond	Mary	John	Mary	John	1	1	throw	over-the-pond
Mary rolled John under the bridge	Mary	John	Mary	John	1	1	roll	under-the-bridge
John was thrown over the pond	-	John	John	-	1	1	throw	over-the-pond
John was rolled under the bridge	-	John	John	-	1	1	roll	under-the-bridge
Mary shrugged the rucksack to the ground	Mary	rucksack	Mary	rucksack	1	1	shrug	to-the-ground
Mary put the spoon on the table	Mary	spoon	Mary	spoon	1	1	-	on-the-table
The spoon was put on the table	-	spoon	spoon	-	1	1	-	on-the-table
The spoon bent	spoon	-	spoon	-	0	1	-	bent
Mary bent the spoon	Mary	spoon	Mary	spoon	1	1	-	bent
The spoon was bent	-	spoon	spoon	-	1	1	-	bent
The spoon went flat	-	spoon	spoon	-	0	1	-	flat
Mary hammered the spoon flat	Mary	spoon	Mary	spoon	1	1	hammer	flat
The spoon was hammered flat	-	spoon	spoon	-	1	1	hammer	flat
John touched the cup	John	cup	John	cup	0	0	touch	-
John picked up the cup	John	cup	John	cup	1	1	pick	up

TABLE I
EXAMPLES ILLUSTRATING THE COVERAGE OF THE NEW WM EVENT MEDIUM

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

[10] D. Dowty, "Thematic proto-roles and argument selection," *Language*, vol. 67, no. 3, pp. 547–619, 1991.

APPENDIX: SUMMARY OF THE COVERAGE OF THE NEW WM EVENT MODEL

Table I illustrates the range of sentence types that can be modelled with our proposed scheme.