When Push Comes to Shove: A Study of the Relation Between Interaction Dynamics and Verb Use

Clayton T. Morrison† (clayton@isi.edu)
Erin N. Cannon‡ (ecannon@psych.umass.edu)
Paul R. Cohen† (cohen@isi.edu)
†Information Sciences Institute; University of Southern California
Marina del Rey, CA 90292 USA
‡Department of Psychology; University of Massachusetts
Amherst, MA 01003 USA

Introduction

A significant portion of our language is devoted to referring to, expressing, and representing the temporally extended dynamics of our world. Following Tomasello (1992), we label the class of words used to refer to such dynamics as "verbs." Cohen (1998) presents a framework for distinguishing interactions involving whole-body objects by considering the dynamics of the before, during (contact), and after phases of interaction. These phases are characterized using dynamic maps to plot various measures of the physical interaction. The *maps-for-verbs* framework proposes that a representation based on the dynamics of before, during and after interaction are a foundation for the semantics of verbs describing physical interactions between objects. Here we report a preliminary study of the use of this framework to predict the verb use of adults describing simple whole-body interactions

The Maps-for-verbs framework

The maps-for-verbs framework proposes that simple interactions between whole bodies can be characterized by the physical dynamics of the interaction. According to the framework, whole-body interactions are naturally divided into three phases: before, during and after contact. Figure 1 depicts these three phases. The interaction of two bodies can be plotted in a two-dimensional space called a *map* (also called a phase portrait or phase diagram). A map portrays the changes in features of or relationships between the two bodies, over time. A given interaction is then described as a trajectory through the map's dynamics space (example trajectories are shown in Figure 1). These maps enable identification of characteristic patterns present in the dynamics of classes of interactions.

Of course, whether the map tells us something useful about the dynamics of the interaction as it relates to identifying or modeling the referent of a verb depends on the features and relations that make up the map's dimensions. Cohen (1998) proposes that the *before* and *after* phases should map relative velocity against the distance between the two bodies. Relative velocity is the difference between the velocity of one body, A, and another, B: $Velocity(A) - Velocity(B)$. Many verbs (e.g., transitive verbs) predicate one body as the "actor" and the other as the "target" (or "subject" or "recipient") of the action. For example, in a scenario involving a PUSH, the actor is the one doing the pushing, and the target is the body being pushed. By convention, the actor is designated as body A and the target is body B. Thus, when relative velocity is positive, the actor's velocity is greater than that of the target; and when relative velocity is negative, the target's velocity is greater than that of the actor. Distance, in turn, is the measure of the distance between the bodies.

The *during* phase is proposed to map between perceived energy-transfer (from the actor to the target), and some other measure, such as time or distance. If energy-transfer is positive, then the actor is imparting to the target more energy than the target originally had; if energy-transfer is negative, then the situation is reverse: the target is imparting more energy to the actor. To measure perceived energy-transfer, we used the simplification of calculating the acceleration of the actor in the direction of the target while in contact. In the original proposal, the second dimension of the during/contact map was a measure of the distance traveled by both bodies away from the initial contact-point. For this work, we instead used the amount of time the bodies were in contact.

Figure 1 depicts a set of labeled trajectories that characterize the component phases of seven interaction types as described by the verbs push, shove, hit, harass, bounce, counter-shove and chase. Using these labels, an interaction can be described as a triple of trajectory labels, indicating the before, during and after characteristic trajectories. For example, $\langle \mathbf{b}, \mathbf{b}, \mathbf{b} \rangle$ describes a *shove.* In a shove interaction, the actor approaches the target at a greater velocity than the target, closing the distance between the two bodies. As it nears the target, the actor slows, decreasing its velocity to match that of the target. Trajectory **b** of the before phase in Figure 1 illustrates these dynamics, showing the decrease in relative velocity, along with decrease in distance. At contact, the relative velocity is near or equal to zero. During the contact phase, the actor rapidly imparts

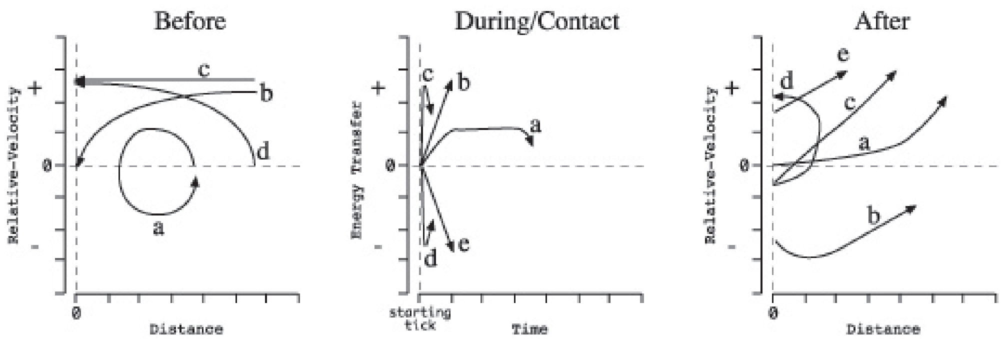

Figure 1: Maps-for-verbs model of the three phases of interaction.

more energy to the target in a short amount of time, as illustrated by **b** of the during/contact phase. And after breaking-off contact with the target, the agent rapidly decreases it's velocity while the target moves at a greater velocity from the energy imparted it (trajectory **b** in the after phase).

A demonstration of the dynamics for shove, based on our simulator (described below), is depicted in Figure 2(b). This map plots the dynamics for a portion of the time between contact phases, beginning with very low relative velocity, as would be expected just after completing the contact phase of a shove (after phase **b**), and ending with a high relative velocity that is ramping down (before phase **b**), as is expected in the period during which a new shove is about to take place, as the actor slows to prepare for contact.

Following this scheme, the remaining six interaction types are characterized by the following triples:

Push $\langle \mathbf{b}, \mathbf{a}, \mathbf{a} \rangle$ – Begins like shove, but at contact relative velocity is near or equal to zero and the actor smoothly imparts more energy to the target; after breaking contact, the agent gradually decreases its velocity.

Hit $\langle \mathbf{c}/\mathbf{d}, \mathbf{c}, \mathbf{c} \rangle$ – May begin with the actor already at high velocity relative to the target or increasing in relative velocity, and thus is characterized by **c** or **d** in the before phase.

Harass $\langle \mathbf{c}/\mathbf{d}, \mathbf{c}, \mathbf{d} \rangle$ – Similar to a hit, except the after-phase involves the actor quickly recovering its speed and moving back toward the target, not allowing the distance between the two to get very large (after phase **d**). Harass highlights that all interactions are not to be viewed only as single movement to contact, but may involve many such movements to contact, one after another, and may even switch between different kinds of contact interactions.

Bounce $\langle \mathbf{c}/\mathbf{d}, \mathbf{d}, \mathbf{e} \rangle$ – Along with counter-shove, bounce involves the target making a more reactive response to the actor's actions. Bounce begins like a hit or harass, but at contact, the target transfers a large amount of energy back to the actor.

Counter-shove $\langle \mathbf{b}/\mathbf{c}/\mathbf{d}, \mathbf{e}, \mathbf{e} \rangle$ – A version of a shove where the target imparts energy to the actor.

Chase $\langle \mathbf{a}, -, - \rangle$ – The agent moves toward the target, closing the distance between the two, but never quite making contact, so the during and after phases are not relevant. This is depicted as the circular trajectory **a** in the before phase.

We used these seven classes of interaction as the basis for a study in which we looked at the frequency of verb usage of adults asked to describe the interaction types after observing them.

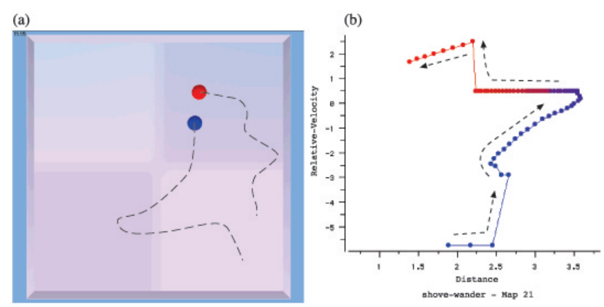

Figure 2: (a) Example of maps-for-verbs simulation running the shove-wander action, as rendered in breve. (Note: dashed lines represent motions of colored patches for demonstration purposes; only the moving color-patches themselves were displayed in the stimuli movies.); (b) Dynamic map plot of shove-wander action before contact, corresponding to the picture in (a) (x-axis = distance between agents, y-axis = relative velocity).

Method

Participants

Forty-four undergraduates (M = 20.5 years old) at the University of Massachusetts participated in this study.

Stimuli

We used *breve 1.4*, an environment for developing realistic multi-body simulations in a three dimensional world with physics (Klein, 2002), to implement a model of the seven interaction classes described in the previous section. The model is rendered as two generic objects (a blue ball for the actor and a red ball for the target) moving on a white background (see Figure 2(a)).

We generated a set of movies based on the rendered interactions. For several of the interaction classes we also varied the behavior of the target object, as follows: the target object, (a) did not move except when contacted ("stationary"), (b) moved independently in a random walk ("wander"), or (c) moved according to billiard ball ballistic physics, based on the force of the collision ("coast"). We generated a total of 17 unique movies.[1]

The 17 movies were recorded and presented on a G3 iMac with 14 inch screen.

Procedure

A total of 18 movies were presented to each participant, with chase being viewed twice. After watching a movie, participants were asked to write down an answers to questions on a sheet of paper given to them by the experimenter. The questions were the same for every movie:

1. What are the balls doing in this movie? (Give your overall impression of what was happening between them, the gist)
2. What is the red ball doing?
3. What is the blue ball doing?
4. Can you think of any words to describe the tone or the mood of the movie? (e.g., the balls are friendly/ not friendly)

The experimenter encouraged participants to write as much as they could to describe the movies.

Measures

In both studies, all the action words and other content words for each trial were extracted and "canonicalized," converting verbs in different tenses or forms (e.g., ending in -ed, -ing, etc.) to a unique form. Also, negation phrases, such as "it's not zooming" or "red didn't move," were also transformed into a single token, e.g., not-zooming and not-moving.

After canonicalization, we kept only the verbs from the content words (a total of 155 verbs). The following 65 verbs are those that were each used by ten or more subjects to describe the movies: advancing, annoying, approaching, attaching, attacking, avoiding, backing, beating, bouncing, bullying, bumping, catching, charging, chasing, circling, coming, controlling, defending, dominating, escaping, fighting, floating, following, forcing, getting, giving, guiding, helping, hitting, kissing, knocking, leading, leaving, letting, looking, losing, nudging, persuing, placing, playing, propelling, pushing, repeating, repelling, resisting, responding, rolling, running, shoving, slamming, slowing, sneaking, standing, standing-ones-ground, staying, stopping, striking, tagging, teasing, touching, traveling, trying, waiting, wanting, winning.

[1] For the bounce and counter-shove interaction types, we only implemented "stationary" and "wander" target behavior, as "coast" would obliterate the effect of the target transferring energy back to the actor. Also, there was only one version of "chase" used, as the target must always be moving away from the actor. Chase was also unique because it was the only instance in which the two balls never contacted each other.

Results and Discussion

Recall that the maps-for-verbs framework hypothesizes that a representation based on the dynamics of before, during and after interactions are a foundation for the semantics of verbs describing physical interactions between objects. If this hypothesis is correct, we would expect the subjects in the preceding experiment to use particular verbs when describing the movies they observed. Furthermore, movies that share the same kind of dynamics in terms of before, during and after phases of interaction should elicit similar groups of verbs. To see whether this was the case, we clustered the 17 movies according the frequency of word usage, where frequency was according to the number of different subjects who used a given word to describe a movie (that is, if five different subjects used the word "approaching" to describe the harass-wander movie, then the frequency recorded was 5). We used hierarchical agglomerative clustering to cluster the movies based on these word frequencies. Figure 3 shows the generated dendrogram tree depicting the results of clustering (ignore for the moment the additional labels and notation to the right).

At first the dendrogram looks disappointing; while there is some structure, it is not clear how to interpret the groupings. However, recall that the movies were generated by behavioral programs, written in breve, that attempt to match the dynamics outlined in Figure 1. The program specifications do not guarantee that the salient perceptual features of before, during and after interaction dynamics will be perspicuous.

To explore this further, we independently observed each movie and chose what we believed to be features that help distinguish movies from oneanother. We came up with a total of five very simple features: whether red (the target of the interaction) looked purposeful before or after contact (*purpose-before*, *purpose-after*) – "purposeful" was in terms of whether red appeared to change its heading on its own; whether red seemed to react to contact (*reactive-during*) – "react" was in terms of whether red appeared to change its behavior based on blue's

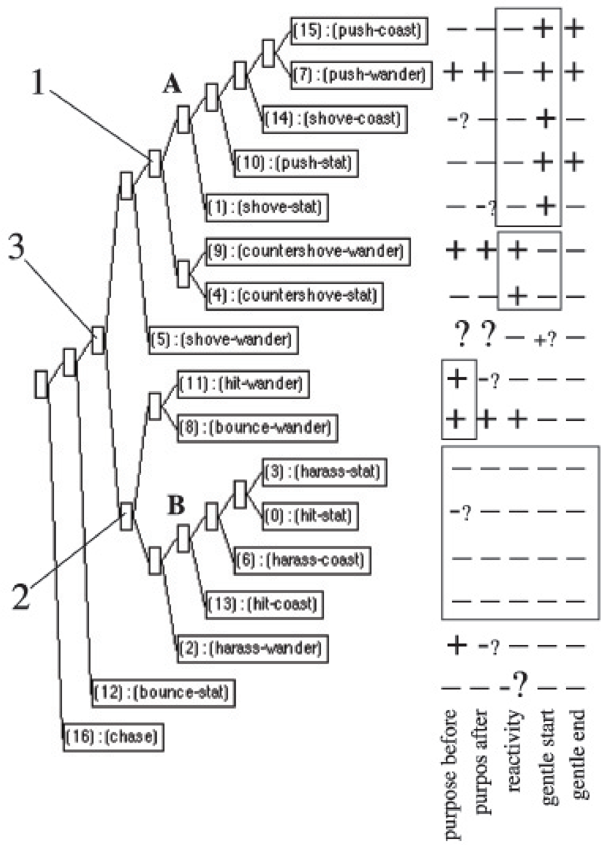

Figure 3: Dendrogram representing clustering of movies based on word usage frequencies, where word usage is based on the number of different subjects who used a given word. The complete set of 155 verbs were used to characterize word usage. The labels inside the leaves of the dendrogram correspond to movie names; the numbers are unique identifiers assigned by the clustering procedure and should be ignored.

contact; and whether the initial or final stages of the contact appeared gentle (*gentle-start*, *gentle-end*).

We then went through each movie and gave a minus (= no) or plus (= yes) depending on whether each feature was present; some cases were uncertain, so we assigned a +? or -?; and some cases were indeterminable (receiving a ?). The features associated with movies are depicted in Figure 4.

We have placed these feature vectors next to the corresponding leaves of the dendrogram in Figure 3. We can now see that there is significant structure to the clusters, based on the similar features that are grouped. The internal node labeled 1 in the dendrogram tree of Figure 3 appears to distinguish between the cluster of movies where red is not reactive to blue's contact while the contact begins gently from movies in which red is reactive and contact does not begin gently. The node labeled 2 in the dendrogram appears to distinguish between whether red looks purposeful before or after interaction (although the placement of harass-wander is problematic; it should be associated with hit-wander and bounce-wander). Finally, the node labeled 3 appears to separate groups of movies that involve gentle starts to interactions or red reactivity from movies that all involve abrupt starts and ends to the contact phase of interaction (except for bounce-wander). We generated a second dendrogram using the more restricted 65 verbs used by 10 or more subjects. The resultant tree is very similar, with the subtrees labeled A and B in Figure 3 being identical.

	purpose before	purpose after	reactive during	gentle start	gentle end
push coast	–	–	–	+	+
push stat	–	–	–	+	+
push wander	+	+	–	+	+
shove coast	-?	–	–	+	–
shove stat	–	-?	–	+	–
shove wander	?	?	–	+?	–
hit coast	–	–	–	–	–
hit stat	-?	–	–	–	–
hit wander	+	-?	–	–	–
harass coast	–	–	–	–	–
harass stat	–	–	–	–	–
harass wander	+	-?	–	–	–
bounce stat	–	–	-?	–	–
bounce wander	+	+	+	–	–
countershove stat	–	–	+	–	–
countershove wander	+	+	+	–	–

Figure 4: Table of simple perceived features of interaction dynamics. (We do not include chase because it is clearly different than the rest)

We believe these results indicate that the dynamical features present in the movies have had an impact on the verbs used by the subjects to describe the movies. Although, to date, we have only tested a subset of the possible interaction types outlined in Figure 1, the data thus far seem to fit the framework of a natural division of interactions into before, during and after interaction phases, and all represent characteristic regions along dimensions of dynamical maps.

References

Cohen, Paul R. (1998). Maps for Verbs. *Proceedings of the Information and Technology Systems Conference, Fifteenth IFIP World Computer Congress* (pp. 21–33).

Klein, J. 2002. breve: a 3D simulation environment for the simulation of decentralized systems and artificial life. *Proceedings of Artificial Life VIII, the 8th International Conference on the Simulation and Synthesis of Living Systems.* http://www.spiderland.org/breve/

Tomasello, Michael. (1992). *First verbs: a case study of early grammatical development.* Cambridge [England]; New York, NY, USA: Cambridge University Press.