# A Framework for Representing Action Meaning in Artificial Systems via Force Dimensions

Paul Hemeren[1] and Peter Gärdenfors[2]

[1] Interaction Lab, School of Humanities and Informatics, University of Skövde, S-541 28 Skövde, Sweden
paul.hemeren@his.se
[2] Lund University Cognitive Science, Lund University, S-221 00 Lund, Sweden
Peter.Gardenfors@lucs.lu.se

**Abstract.** General (human) intelligence critically includes understanding human action, both action production and action recognition. Human actions also convey social signals that allow us to predict the actions of others (intent) as well as the physical and social consequences of our actions. What's more, we are able to talk about what we (and others) are doing. We present a framework for action recognition and communication that is based on access to the force dimensions that constrain human actions. The central idea here is that forces and force patterns constitute vectors in conceptual spaces that can represent actions and events. We conclude by pointing to the consequences of this view for how artificial systems could be made to understand and communicate about actions.

**Keywords:** Intentions, actions, conceptual spaces, force, action representation, concepts, categorization, events.

## 1 Introduction

A critical aspect of human intelligence is reading the intentions of others, being able to understand their actions and the extent to which future behavior is determined by current behavior. This speaks to the predictive role in intelligent behavior. Much recent evidence (for a review see [1]) has demonstrated the role of human movement in understanding the social interaction between humans. Human movement can carry social information about the intentional state and future motor states of humans. Intelligent artificial systems that can use/exploit this information will have an advantage over systems that do not have access to this information. Intelligent artificial systems will need to both recognize human intention-governed action as well as produce actions that can be understood by humans. This is what humans seem to do, i.e., read each other's intentions and act accordingly.

Individual actions have specific activations of muscles which give rise to specific kinematic patterns. Being able to generalize beyond these situation-specific activations will allow us to act and understand other individuals across varying contexts. For example, throwing a stone can be done in a number of ways, and the exact kinematics (as well as dynamics) will vary depending on the individual and the object being thrown. Despite these contextual differences, we can easily see and describe what others are doing on a level that generalizes over the contextual differences. In this case, perception and language will both rely on access to an action representation hierarchy. Similar to objects, we generalize from specific examples of actions to broad action categories.

In this article we will outline a unifying framework connecting action meaning with cognitive structures for action representation and categorization. We argue that this framework captures critical aspects of human action representation that can used by artificial systems to appropriately understand human actions. The central idea is that perceptions of forces and force patterns are appropriate basic elements in generating action categories. We also present evidence for the hierarchical organization of action categories and demonstrate the relation between this organization and our ability to communicate about actions on different hierarchical levels.

## 2 Representing actions as conceptual spaces: sensorimotor grounding and force dimensions

There is considerable evidence for the sensorimotor grounding of action concepts [2]. These concepts allow us to create different categories of actions on the basis of patterns of similarity and movement. Running, walking, throwing, waving, etc., have characteristic kinematic patterns, and these patterns affect our perception of what others are doing as well as the words we use to communicate about human actions [3-4]. Recent results [5] also show the significant relationship between action understanding and human movement kinematics for hand and arm actions.

The idea here is that successful interpretation of human actions as a basis for social interaction requires grounding action understanding the physical constraints that govern movement of the human body. This includes both internal and external constraints. Internal constraints are for example the degrees of freedom that a human body has and the forces that it has to affect objects in the environment. These forces include muscle strength as well as the ways in which the body can move in order to exploit the available muscle strength. The external constraints are the physical properties in the surrounding environment.

Another source of information that can be used to recognize human actions is the kinematics associated with human body movement, i.e., motion as such. There is a relationship between kinematics and dynamics in the sense that the dynamics (forces) will to some extent determine the kinematic patterns associated with different human actions.

When one perceives an action, one does not just see the movement; one also extracts the forces that control different kinds of motion. [6] formulates this as the principle of *kinematic specification of dynamics*, which says that the kinematics of a movement contain sufficient information to identify the underlying dynamic force patterns. Our proposal is that, by adding forces, one obtains the basic tools for analyzing the dynamic properties of actions. The language of vectors will be of great representational convenience here.

## 3 Actions as convex regions

[7] argues that properties can be represented by convex regions of conceptual spaces. For example, the property of being red is represented by a convex region of the three-dimensional color space. Certain sets of dimensions are integral, for example the color dimensions, in the sense that one cannot assign an object a value on one dimension without giving it a value on the other(s) [8]. We define a domain as a set of integral dimensions.

A concept – in the most general sense – can then be defined as a bundle of properties that also contains information about how the different properties are correlated. For example, the concept of an apple has properties that correspond to regions of the color domain, the shape domain, the taste domain, the nutrition domain, etc. Correspondingly, object concepts can be represented as a complex of properties from a number of domains: that is, bundles of properties.

Action categories can be represented as convex regions in a conceptual space. Accordingly, a convex region is characterized by the criterion that for every pair of points $v_1$ and $v_2$ in the region all points in between $v_1$ and $v_2$ are also in the region. The implication of this notion of convexity when applied to action categories is that if two actions are categorized as exemplars of the same category, and they occupy separate points in a convex region, then any action exemplar occupying a space between then will be categorized as belonging to the same category.

It is important to note that this framework takes the context of categorization into account by stipulating that the dimensions of actions determine the basis for assigning properties to actions as well as determining the relations among the properties. In this sense, different contexts, perhaps defined by different goals or other situational factors, will lead to the use of different dimensions and thereby different regions of convexity.

For many actions – for example moving and lifting – a single force vector may be sufficient, but for others – such as walking and swimming – a complex of forces is involved. We therefore define an action as a pattern of forces since several force vectors are interacting (by analogy with the system of differential equations in [9]).

This framework views the action spaces as geometrical structures, and as such we can view objects/actions as being psychologically *closer* (more similar) or *further* from one another (less similar) in vector space. To identify the structure of the action space, one should investigate similarities between actions. For example, walking is more similar to running than to throwing. An action category can then, in the same way as with other kinds of categories, be characterized as a convex region in a space of force vectors or force patterns. [10-11] present some further empirical evidence that supports this definition of action categories.

In a way similar to object categories, results from psycholinguistic experiments show that action categories also appear to have a hierarchical taxonomic organization [12]. It should be noted that the above definition of conceptual spaces immediately generates a model of hierarchies of action categories. The idea is that if a region representing action category A is a subregion of the region representing category B, then A is a subcategory of B. For example, the force patterns corresponding to the verbs march, stride, strut, saunter, tread, etc., can all be seen as subregions of the force patterns that describe walk.

### 3.1 Action prototypes

If force dimensions are central aspects of actions and action representation that serve perception and production, then we would expect these dimensions to play a critical role in the structure/metric of conceptual spaces for actions. The results from [13] provide support for this view of a metric representation of action categories. They created morphs between four action prototypes for *running*, *walking*, *limping* and *marching*. This resulted in a space of many different action morphs that were combinations of for example *running and walking* or *running and marching*, etc. Human subjects then provided ratings regarding the naturalness of the action morphs that existed somewhere between the different prototypes. These naturalness ratings could be reliably predicted on the basis of knowing what (force) dimensions were used to create the morphs from the original action prototypes. In this case, the psychological data could be mapped onto the action space created by different linear combinations of each prototype. This shows that the convexity of the action space represented by action morphs reflected the rating behavior of the subjects.

Action perception and representation appear to be structured around a kind of motor template (prototype) that generates a force pattern which best represents a given category of action [3]. Action concepts therefore contain information about prototypical spatiotemporal patterns of human movement.

Access to an action prototype is also consistent with data from human categorization results. Results from typicality and action verification studies using point-light displays of biological motion [12] indicate that the time to verify the category membership of different kinds of walking, kicking, throwing and waving are highly inversely correlated with judgments of typicality for those same actions. An important factor here is that this occurs as a result of reading a category label (WALKING) and then viewing different point-light displays of walking instances. There is considerable subject agreement about which actions are prototypical for the different action categories and which action exemplars are atypical, or poor examples of the action category.

## 4 The two-vector model of events

The analysis of actions can be used as a basis for modeling events. The model presents events as complex structures that build on other conceptual spaces: in particular, the action space. The central idea is that an event can be cognitively represented as a mapping between two types of vectors:

- *The two-vector condition*: A representation of an event contains at least two vectors and at least one object – a result vector representing a change in properties of the object and a force vector that represents the cause of the change.

The structure of the event is determined by the mapping from force vector to result vector. We call the central object of an event the patient.

As a simple example of the model, consider the event of John moving a book. The force vector is generated by an agent: John. The result vector is a change in the location of a book and thus a change in the book's properties. The outcome depends on the properties of the patient as well as other aspects of the surrounding world: e.g., friction. Even though prototypical event representations also contain an agent, some event representations need not involve an agent: e.g., in cases of falling, drowning, dying, growing, and raining.

The vectorial representation of forces provides a natural spatialization of causation that unifies the model with other applications of conceptual spaces. In the limiting case when the result vector is the identity vector (with zero length), the event is a state. However, states can be maintained by balancing forces and counterforces: for example, when a prop prevents a wall from falling.

Notice that since force and result vectors can form categories – as convex spaces of mappings – a natural extension is that events also form categories, as mappings between action categories and change categories. For example the set of all force vectors involved in moving a book is naturally convex, and so is the set of all paths (change vectors) of moving the book to the desk.

The proposed model allows one to represent events at different levels of generality. There are subcategories of events, just as for object and action categories. For example, *pushing a door open* is that subcategory of *pushing a door*, where the force vector exceeds the counterforce of the patient. *Pushing a door but failing to open it* is another subcategory, where the counterforce annihilates the force vector.

A limiting case of the event model, expressed linguistically by intransitive constructions such as “Victoria is walking” and “Oscar is jumping,” is when the patient is identical to the agent. In these cases, the agent exerts a force on him/her/it/self: in other words, the agent modifies its own position in its space of properties.

## 5 Communicating actions via verbs

The two-vector model of events has immediate consequences for how verbs can be learned and used by robots. The topic of verb learning in robot is currently studied by several groups [14-16]. These attempts, however, focus on result verbs. For example, [14] used seven behavior categories: push-left, push-right, place-left, place-right, push-forward, place-forward and lift. (Even though “push” is a manner verb, it is used in the “move” meaning in this context). Their algorithms for learning the verb meanings are based on “affordance relations” between entities, behaviors and effects. In terms of the semantic model presented here, entities correspond to patients, behaviors to force vectors and effects to result vectors. [14] then present vector-based models for how an iCub robot can extract prototypical effect (result) vectors for the seven categories above.

We propose that this methodology be extended also to manner verbs. For example to be able to distinguish between “push” and “hit”, the robot should calculate and categorize the force vectors in the actions. If it is observing another agent pushing or hitting, the force vectors can be extracted from the second derivatives of the kinematics of the movements (for example exploiting the methods of [3]. We know from [17] that the human brain is extremely efficient in this.

Then the robot must learn the mapping from force vectors to result vectors. Hitting a ball will have different consequences from gently pushing it. Such an associative mapping can be extracted from a combination of observing the force and result vectors of other agents interacting with objects and learning from the robots own interactions with objects and their results.

The framework can also be applied to verb meaning by explaining similarities between verbs, by building on the distances between the underlying vectors. The fact that the meaning of walk is more similar to that of jog than that of jump can be explained by the fact that the force patterns representing walking are more similar to those for jogging than those for jumping. Although we have not presented the details of the similarities of the actions involved, these can be worked out systematically from the proposed vectorial representation of actions.

In a parallel way, the theory explains the general pattern of the sub-categorizations of verbs: For example, the force patterns corresponding to the verbs march, stride, strut, saunter, tread, etc., can all be seen as subsets (more precisely, sub-regions) of the force patterns that describe walk. The inference from e.g. “Oscar is marching” to “Oscar is walking” follows immediately from this inclusion of regions within one another. No previous theory of verb semantics seems to account for these two central properties.

Finally, [18] distinguish between manner verbs and result verbs – where “manner verbs specify as part of their meaning a manner of carrying out an action, while result verbs specify the coming about of a result state” [19]. The single domain constraint provides an immediate explanation of this distinction, mapping manner verbs onto the force vector and the result verbs on the result vector.

### 5.1 Summary of the semantics

The building blocks for the semantics of verbs are two extensions of the theory of conceptual spaces: (i) a model of actions as patterns of forces, and (ii) a model of events as couplings of force vectors (patterns) and result vectors associated to a patient space.

Using these models, the main semantic thesis is that verbs refer to convex regions defined by a single semantic domain (as do adjectives). Together with the framework of conceptual spaces, this approach explains many features of the semantics of verbs. By focusing on vector representations, one obtains a strong tool for systematizing linguistic data.

## 6 Conclusions and applications

The framework put forth here reflects central findings regarding human action perception and production. Conceptual spaces, and specifically force vectors and force patterns, can be used to model our understanding and communication about the actions of the other humans. The ability to understand and communicate about our own actions and the actions of other is a fundamental aspect of our daily activity. The main cognitive elements are action representations and the main linguistic elements associated with actions are verbs.

A central question is therefore how speakers construe the mapping between actions and action representations either a concepts or linguistically as verbs. Within in the past 10 years, increasing evidence [20-21] indicates a close mapping between sensorimotor activity and verb meaning. Furthermore, force and kinematic information about the motion of other bodies seems to be a shared basis for verb understanding. For example, recent results [22] show that information about grip force is encoded in the meaning of manual action verbs.

Regarding learning, the robot must learn the mapping from force vectors to result vectors. Hitting a ball will have different consequences from gently pushing it. Such an associative mapping can be extracted from a combination of observing the force and result vectors of other agents interacting with objects and learning from the robots own interactions with objects and their results.

A complicating factor is that various *contextual* factors may influence the connection between the force and the result vector. For example, pushing an object on ice may lead to different results than pushing the same object in the same way on a lawn. This means that the learning involves a mapping from the three factors: force vector, object, and context to the result vector. Extracting such a mapping is not an easy task, even in a simplified environment. However, by using clustering techniques or verbal input to the robot, the aim should be to learn a mapping from categories of force vectors and categories of objects to categories of result vectors that can also take some relevant contexts into account.

Implementing such a learning mechanism in an artificial system is a sizeable task. However, this is the kind of mapping that children learn during their first years [21]. By manipulating objects in different ways and in different circumstances, they learn about the consequences of their actions. The learning is scaffolded by the language learning that is going on at the same time. We believe that the two-vector model of events that has been presented here is a powerful tool for implementing the learning of action meaning in artificial systems.

## References


1. Pavlova, M.A.: Biological Motion Processing as a Hallmark of Social Cognition. Cere. Cort. doi:10.1093/cercor/bhr156 (2011)
2. Shiffrar, M.: Embodied Motion Perception: Psychophysical Studies of the Factors Defining Visual Sensitivity to Self- and Other-generated Actions. In: Klatzky, R. L., MacWhinney, B., Behrmann, M. (eds.) Embodiment, Ego-Space, and Action, pp. 113-143. New York: Psychology Press. (2008)
3. Giese, M., Thornton, I., Edelman, S.: Metrics of the perception of body movement. J. of Vis. 8(9) 1-18 (2008)
4. Malt, B., Gennari, S., Imai, M., Ameel, E., Tsuda, N., Majid, A.: Talking about Walking. Psych. Sci. 19(3), 232-240 (2008)
5. Hemeren, P. E., Thill, S.: Deriving Motor Primitives through Action Segmentation. Front. in Psych. doi: 10.3389/fpsyg.2010.00243 (2011)
6. Runesson, S.: Perception of Biological Motion: The KSD-principle and the Implications of a Distal Versus Proximal Approach. In: Jansson, G., Bergström, S.-S., Epstein, W. (eds.) Perceiving Events and Objects, pp. 383-405. Hillsdale, NJ: Lawrence Erlbaum Associates (1994)
7. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. Cambridge, MA: MIT Press. (2000)

8. Melara, R. D.: The Concept of Perceptual Similarity: From Psychophysics to Cognitive Psychology. In: Algom, D. (ed.) Psychophysical Approaches to Cognition, pp. 303-388. Amsterdam: Elsevier (1992)
9. Marr, D., Vaina, L.: Representation and the Recognition of the Movement of Shapes. Proc. R. Soc. Lond. 24, 501-524. (1982)
10. Gärdenfors, P.: Representing Actions and Functional Properties in Conceptual Spaces. In: Ziemke, T., Zlatev, J., Frank, R.M. (eds.) Body, Language and Mind, Volume 1: Embodiment, pp. 176-195. Berlin: Mouton de Gruyter (2007)
11. Gärdenfors, P., Warglien, M.: Using Conceptual Spaces to Model Actions and Events. J. of Sem. doi:10.1093/jos/ffa007 (2012)
12. Hemeren, P. E.: Mind in Action. Lund University Cognitive Studies 140. Lund. (2008)
13. Giese, M.A., Lappe, M.: Measurement of Generalization Fields for the Recognition of Biological Motion. Vis. Res. 38, 1847-1858 (2002)
14. Kalkan, S., Dag, N., Yürüten, O., Borghi, A. M. & Sahin, E. (to appear). Verb concepts from affordances. (to appear)
15. Lallee, S., Madden, C., Hoen, M., Dominey, P. F.: Linking Language with Embodied and Teleological Representations of Action for Humanoid Cognition. Front. in Neurorobot. 4, Article 8 (2010)
16. Tikhanoff, V., Cangelosi, A., Metta, G.: Integration of Speech and action in Humanoid Robots: iCub Simulation Experiments. IEEE Trans. Aut. Mental Dev. (3)1, 17-29 (2011)
17. Johansson, G.: Visual perception of biological motion and a model for its analysis. Perception & Psychophysics, 14(2), 201-211 (1973)
18. Levin, B., Rappaport Hovav, M.: Lexicalised Meaning and Manner/Result Complementarity. In: Arsenijevic, B., Gehrke, B., Marin, R. (eds.) Subatomic Semantics of Event Predicates. Dordrecht: Springer. (to appear)
19. Rappaport Hovav, M., Levin, B. Reflections on Manner/Result Complementarity. In: Rappaport Hovav, M., Doron, D., Sichel, I. (eds.) Lexical semantics, Syntax, and Event Structure, pp. 21-38. Oxford: Oxford University Press (2010)
20. Pulvermüller, F.: Brain Mechanisms Linking Language and Action. Nat. Rev. Neuro. 6, 576-582 (2005)
21. van Dam, W. O., Rueschemeyer, S-A., Bekkering, H.: How Specifically are Action Verbs Represented in the Neural Motor System: An fMRI Study. NeuroImage, 53, 1318-1325 (2010)
22. Frak, V., Nazir, T., Goyette, M., Cohen, H., Jeannerod, M.: Grip Force is Part of the Semantic Representation of Manual Action Verbs. PLoS ONE 5(3): e9728. doi: 10.1371/journal.pone.0009728 (2010)
23. Thelen, E., Smith, L.B.: A Dynamic Systems Approach to the Development of Cognition and Action. Cambridge, MA: MIT press (1994)