Signs, Deixis, and the Emergence of Scientific Explanation

Wolff-Michael Roth and Daniel V. Lawless

Introduction

While many scholars consider gestures an integral (rather than epiphenomenal) feature of communication, most researchers regard gesturing as psychological or linguistic phenomena. Consequently, few models have been developed to study gestures from a semiotic perspective. With this in mind, we propose a semiotic model for the purpose of studying the relationship between speech and gesture during the emergence of scientific explanations. We draw on seminal work in phyto- and zoosemiotics (e.g., Krampen 1986; Sebeok 1986) to conceive of the objective structures of the lived world as signs: this then allows us to track signs and their relations in the three modalities of percept, gesture, and speech. Furthermore, rather than focusing on a classification of signs, we follow Eco (1976) and study the modes of sign production and how their changes are constitutive of the ontogenesis of scientific language. In this way, we contribute not just to the debate between the two dominant theories on gesture-utterance relationship (e.g., Butterworth and Hadar 1989; McNeill 1985) but add a new dimension heretofore not considered in this debate—developmental issues in the emergence of language.

In addition to the semiotic design, our approach is unique from other research on gestures for a number of reasons. First, research in this area has generally focused on communicative situations where individuals talk about something familiar (or a story they heard earlier in the paradigm) and typically as competent speakers. Consequently,
we know little about gestures, language, and their relationship when individuals try to make sense in an unfamiliar domain, which is often the case with new scientific phenomena. Second, past research has generally focused on gesture-utterance relations at a single point in time. Consequently, there is a dearth of research on the long-term changes in communicative patterns as individuals become increasingly familiar with a topic. Third, much of the existing research is concerned with gestures people use as they are talking about non-present events (for exceptions see studies in anthropological linguistics [e.g., Haviland 1993; Levinson 1997; Widlok 1997]). As a consequence, the indexical ground of the communicative situation remains an under-researched phenomenon (Goodwin and Duranti 1992). Indeed, ‘there has been relatively little in-depth description of actual usage [of deixis], and available descriptive frameworks are partial and relatively coarse’ (Hanks 1992: 42).

In this study, we simultaneously address all three points. Our aim is to provide better descriptions of the semantic and pragmatic mechanisms of deixis in situations where participants face new phenomena for which they are to develop observational and theoretical language. Over the past decade, we videotaped students in school science laboratories as they experienced and were learning about phenomena heretofore unknown to them. In each study, the communication of particular students were studied over long periods of time (2 to 4 month) and as they developed from not knowing some phenomenon to the point that they competently communicated about it. We recognized only recently the important role of gestures (symbolic movements) seem to play in the emergence of scientific language from laboratory activities that require the manipulation
ergotic movements) and sensing (epistemic movements) of entities in the world (e.g., Roth 1996; 1999a; 1999b; in press).

Our contribution to the understanding of developmental issues has at least three components: (1) we contribute to the understanding of long term developments in the evolution of explanations beginning with ‘raw’ experience; (2) we articulate the evolution of figure-ground distinctions which, subsequently, leads to the emergence of gestures and talk, and ultimately, to scientific explanations; and (3) our work contributes an understanding of the evolution of socially acceptable ways of seeing and talking about phenomena.

We chart the following itinerary. First, we review existing theories and recent work on the evolution of gesture-speech relations. We then propose a new semiotically grounded framework for understanding evolutionary processes in the genesis of perception, gesture, and speech when people begin to see and know about new phenomena. In the section ‘From Raw Stimulus to Science’, we provide exemplary data from our research in school science classrooms that show how students’ talk evolves from rather incoherent to socially acceptable talk about given phenomena. (Because we are interested in the emergence and evolution of explanations per se, we are less concerned whether a particular explanation coincides with some standard explanation.) Finally, we discuss three issues related to our findings: the multi-modality of communication, the emergence of scientific language, and the notion of iconicity in relation to gestures.
Previous Research on Gesture Speech Relations

Existing Theories of the Gesture-Speech Relationship

In the current scientific literature, there are two main theories about the nature and function of hand gestures that accompany speech (Beattie and Shovelton 1999). One of these theories suggests that gestures do not convey any semantic information over and above what the linguistic utterances they accompany convey (Butterworth and Hadar 1989; Hadar and Butterworth 1997). In this theory, gestures are epiphenomenal to linguistic expression, which itself arises from preexisting semantic models of what is being talked about. When the speaker searches for an appropriate word, there will be pauses that increase with the unfamiliarity of the word (and possibly topical domain). In such cases, utterances are delayed in respect to the corresponding gestures. There appear to be ‘four distinct, but not necessarily mutually exclusive functional roles of gestures’ (Butterworth and Hadar 1989: 173). First, gestures may be a way of signaling to the listener that the speaker has not completed his turn; therefore, they act as an interruption suppression signal. Second, gestures may increase the overall neuronal activation potential, which helps to prime a word’s firing potential more quickly when the build-up of activation in the searched-for item is too slow. Third, the production of an iconic gesture may facilitate word search by exploiting a different route to the phonological lexicon. Butterworth and Hadar suggest that this process would require a recoding of the semantic model into a perceptual model. Fourth, gestures may leak through although the corresponding selected words have been censored and suppressed for some reason before they can be uttered. In summary, Butterworth and Hadar claim that iconic gestures that
occur in speech are epiphenomenal and can be attributed to different aspects of lexical search, including lexical retrieval (Figure 1a).

In contrast, McNeill (1985; 1992) holds that gestures and speech share a computational stage and are therefore part of the same psychological structure. In Figure 1b, which contains additional features to be discussed below, we express this shared stage as the ‘semantic model’. Because the semantic model drives both utterances and gestures, they constitute in fact alternate avenues for expressing semantic content. Alternative avenues are particularly important when there are gaps in the lexicon (McNeill 1992) or when gestures, because they are better at expressing topologic features, are a more appropriate expressive modality (Lemke 1999). To illustrate, McNeill (1992: 129) uses the example of a speaker who utters ‘and he steps on the part [wh]ere [th]e/[str]eet [car’s] connecting’ and accompanies this by a ‘flat open palm of left hand contacts upward extended index finger of right hand several times’. McNeill suggests that the gesture is iconic to the repeated up/down bouncing of the protagonist’s foot on the trolley connector. Here, the gesture fills a logical gap by covering a ‘situation in which there is the movement of a particular shape of thing’ (1992: 129) and therefore expresses both manner and cause of movement not normally covered by English verbs.

We find McNeill’s work particularly appealing because it is commensurable with a well-known theory of sign production (Eco 1976), with motor theories of language origin (e.g., Bates in press; Lock 1997), and with the notion of perceptual symbol systems
First, Eco suggested that the model in Figure 1b covers most situations of sign production. Here, conceptual processes extract particular features from a sea of stimuli hitting the sensory surface, which leads to a perceptual model. Through an abstraction, a semantic model is derived which itself gives rise to expressions based on similitude. (Because the notion of iconicity is fraught with many problems, a number of semioticians have suggested to abandon or considerably redefined it [e.g., Eco 1976; Fenk 1997].) Here, though Eco did not write about them, we consider gestures as an expressive means that is consistent with his model. Our own research has shown that there is a high degree of similitude between perceptual features of laboratory investigations and the gestures students use to describe and explain them (Roth in press). In fact, we also showed that in students’ communicative efforts gestures appear (sometimes much) earlier than the corresponding lexical items.

Second, McNeill’s theory is also compelling because it sits well with motor theories of language origin. In these theories (e.g., Bates in press; Lock 1997), gestures are said to have arisen as a new way of making use of neuronal assemblies that have evolved in response to humans’ movements in and manipulations of aspects of the world. As a next step in the communicative abilities of humans, language has evolved by making use of already-existing neuronal assemblies responsible for movement and gesture. This, therefore, suggests a phylogenetic pathway (dotted in Figure 1b) in which the independent sets of expressive units (Figure 1b) have arisen after and on the basis of previously existing (iconic) gestures. Our own work suggests that (ergotic, epistemic) hand movements enacted by students during their laboratory activities facilitate the
emergence of gestures (iconic, symbolic movements) and the subsequent emergence of language (Roth 1999b).

Gesture-Speech Relations during the Ontogenesis of Scientific Language

Research on gesture-speech relations is generally based on the assumption that the sign-producing speaker knows the domain s/he is talking about. This assumption is embodied in the theoretical models depicted in Figure 1. In both theories, there exists a semantic model that drives both speech and gesture production. However, gesture and speech are not always consistent. Goldin-Meadow and her colleagues (1992; 1993) reported discrepancies between gestures and speech when children were in transitional states of their understanding. (Here, we are not considering those situations covered in the Butterworth and Hadar theory as a ‘leaking through’ of expression in the gestural modality, which was successfully suppressed in the linguistic modality.) These authors suggest that during transitional states in cognitive development, the gestures already depict new understandings, even though students’ linguistic competencies to express the content is not yet developed. Thus, the gestural and verbal modalities express different understandings. Our research also provided evidence that when students are asked to make sense in a new topical domain, the gestural modality appears to run ahead in the evolution of new expressive capabilities seemingly indicating the existence of new semantic models (Roth 1999b; in-press). In our dilemma of modeling this situation, we found help in Eco’s work on sign-production in situations where neither content nor expressive continua are shaped.

))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))) Insert figure 2 about here )))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))]}
Eco (1976) suggested that rather than driving iconic and verbal expressions (representations), semantic models are the outcome of processes where the sign-producing individual has to organize not only the expressive means (language, gesture, and other [written] representations) but also develop a structure for the content. Adapting Eco’s model for our purposes, we arrive at Figure 2. Here, experimenting individuals (students, scientists) manipulate objects to foreground events and create phenomena with which they were not familiar. However, during the initial stages of an investigation, the individual does not know yet how to structure the perceptual field into figure and ground. That is, they cannot extract those features that will lead to robust and viable observational and theoretical descriptions. As the individual investigates, the perceptual model, gestures, and language mutually-adjust (see double arrows) until she arrives at a stage where the three modalities are consistent. In short, she gradually constructs a semantic model.

In this way, our model is suited to account for the discovery work of Faraday (Gooding 1990) and the Italian physicist Malpurgo’s search for quarks (Pickering 1995). Both studies describe in great detail the processes by means of which the scientists manipulated that part of the continuum (‘transformation’ in Figure 2) for which they developed, as an outcome of the investigative process, a means of observational and theoretical descriptions (i.e., a semantic model). Our research on learning physics in school science laboratories shows the same kind of underlying structure. Initially, students do not know what is relevant in some perceptual field; they often disagree drastically about what was seen in an experiment or demonstration (Roth et al. 1997a;
1997b). As we show below, in the course of investigating and in attempting to construct (verbal and gestural) observational and theoretical descriptions, students develop ways of structuring the perceptual field and, consistent with it, descriptions and explanations. Importantly, our research documented structural similarities between gestures (symbolic movements) and manipulations (ergotic, epistemic movements), suggesting that the former may have arisen from the latter (Roth 1999a; 1999b; in press). Furthermore, because both perceptual and gestural are encoded by the brain in terms of the same or interconnected sensori-motor neuron assemblies (Barsalou 1999), gestures may be an expressive means that runs ahead in ontogenetic development. In the diagram, we captured this by means of transformations (arrows in Fig. 2) based on similitude.

**Framework for Studying Evolutionary Trends in Gesture-Speech Relations**

It is now well accepted in many disciplines that perception is theory-laden (including ‘common sense’); that is, perception and conception are mutually constitutive. Our own research showed that students not only see physics demonstrations in ways that drastically differ from the teacher but that these perceptions are consistent with their ‘folk’ (i.e., untutored) explanations of what they are seeing (e.g., Roth et al. 1997b). There was a very high correlation between what students expected to see and what they subsequently saw (e.g., whether something moved or not). (Whereas this observation is expected from the perspective of a biosemiotician [e.g., Hoffmeyer 1997; Sharov 1998; Uexküll 1973], this is quite a revolutionary approach within education or cognitive science.) The question then poses itself how science students, for example, can learn to see and explain new phenomena in terms of new theoretical language. Such an evolution
requires a bootstrapping process to get out of the loop that reifies perception and explanation.

Interaction participants frequently do not make reference in talk or gesture to that which is perceptually available to everyone. Consequently, it became necessary to develop a framework that includes the perceptual ground as one modality of communication. Furthermore, the framework would also have to allow for the description of the evolution from not seeing and knowing a phenomenon to the presence of a fully fledged, socially negotiated and accepted explanation of the phenomenon now perceptually available to the participants. These two parts of our framework are outlined in the following subsections.

Modalities of Communication

In everyday life, we use many different modalities to communicate. In addition to the commonly analyzed speech acts and gestures, these modalities also include environmental structuration (e.g., appropriated topography) and artifactual markings (materials) that figure a body’s topologically defined ground (Preziosi 1986). For our purposes, we found it helpful to distinguish three modalities of communication, corresponding to an analysis of signs at three levels: percept (subjective environmental structuration), gesture, and utterance. To elaborate these concepts and our form of analysis, consider the following excerpt from a lesson in which four students sit around a computer, using a computer-based Newtonian microworld (see appendix for a description of the microworld and the underlying physics). At this point in time, the conversation between the students is about force. At the very moment that student J uses the cursor to
manipulate one of two arrows, «velocity», M (who is partially seen in the video from behind) begins to speak (Figure 3).  

As M utters the first part of his contribution (italic indicates verbal emphasis), his arm begins to move upward so that, when he uttered ‘this’, his hand was about 20 centimeters from the screen, index finger pointing upward parallel to «velocity». Here, the computer monitor constitutes the indexical ground. Though participants operate under the assumption that the talk is about objects and events on the monitor, they cannot know what M will say as he begins. As his expression unfolds, the indexical ground splits into figure (that which the talk is about) and the ground (the field from which the figure becomes salient). His listening peers have to construct what M talks about by drawing on signs at the three levels.

At the level of the utterance, we learn that ‘that is not the force’ but ‘this one is direction’. We begin our analysis with the second part. ‘This’ is an instance of verbal deixis, a way of indicating something in the proximity. It coincides with both the right hand in a ‘pointer’ configuration (Eco 1976) and a particular configuration on the monitor (Figure 4). It is this pointer configuration that makes both the arrow and the finger an aspect of the ground from which the sign function emerges that the two take with respect to each other. As indicated in our figure, something in the ground motivates the configuration and spatial orientation of the hand. At the same time, the hand motivates the perceptual processes to structure the field stimuli in a particular way. If the
three modalities are assumed to be consistent with each other, then ‘this’ provides for the coordination across the three modalities, while both finger and vertical arrow are signs that serve as respective interpretants to each other in respect to some entity here identified as ‘direction’. (For physicists, the arrow stands for «velocity».)

In the present episode, deixis was enacted verbally. There are many instances in our database where verbal indexicals are accompanied by pointing or even touching gestures. In this case, the gestural modality also enacts the coordination across modalities, whereas the nature of the entity only appears in the perceptual modality. Here, the entity, named and indexed by the gesture and word, serves the same communicative function that an utterance such as ‘the arrow that points upward stands for direction’ would have served. This may be more evident from an example later in the students’ development when they make statements such as ‘velocity is changed by a force’ which operates independent of any specific investigation they conducted on-screen.

This brief example helps to illustrate some of the central aspects of our analysis. To start, we chose a holistic unit of analysis by including relevant (perceptual) aspects of the physical environment as signs that operate together with what are usually considered the constitutive embodiments of signing processes (words, gestures). We include the percepts (salient and appropriated environmental structures) in our analysis for three reasons.

First, because interpretation can be understood as defining one portion of the continuum to serve as a vehicle for another portion of the continuum ‘derived from its
global segmentation by its content’ (Eco 1984: 44). As such, there exists a symmetric relation between an object such as a tree and a corresponding sign-vehicle for a cultural form also signified by the word ‘tree’. But the percept of tree is afforded through our senses. Percepts can be regarded as the perceptual parallel to the interpretants in Peircean semiotics (e.g., Moriarty 1996). This is because ‘every object, event or act has stimulus value for the members of a society only insofar as it is an iconic sign signifying some corresponding form in their culture’ (Eco 1976: 165, citing Goodenough).

The second reason for including percepts in our analysis is rooted in neuro-psychological considerations. At the level of the brain, modes of perceptual symbol systems exist that differ from the linguistic modality, but are similar to the sensori-motor modalities responsible for moving about and manipulating objects (Barsalou 1999; Decety and Grèzes 1999). The individual therefore responds to signs that are visually coded or referenced but do not exist in other, verbal modes and are therefore absent in traditional accounts of cognition.

Finally, percepts need to be included in the analysis because copresent speakers often do not talk about, point to, or represent in gesture those entities that ‘go without saying’ (and, equivalently, without gesturing). Percepts are the significant and signifying structures in the (subjective) Umwelt that speakers assume to be shared unless communicative breakdown shows that this is not the case.

From a cognitive perspective perhaps the strongest evidence against the common assumption that the things talked about or acted on need to be represented in mind comes from cognitive science and artificial intelligence. Thus, the performance of highly competent Tetris players, short-order cooks, and others could only be accounted for when
cognition is distributed across the environment—that is, when objects in the world represent themselves but are not mirrored as mental concepts (e.g., Agre and Horswill 1997; Kirsh 1995). In short, perceptions constitute a modality that in many instances allows much more complex and different cognition than verbal and amodal modes would ever allow (Barsalou 1999).

To sum up, in our analyses we account for the salient perceptual aspects of the indexical ground as signs that stand for themselves as part of our unit of analysis. Each modality—percept, gestures, and speech—has a ‘horizontal’ component, contributing to unfolding the narrative. Each modality also has a ‘vertical’ component (gestural, verbal deixis) that serves to coordinate the narratives in the different modalities with each other.

**Modes of Sign Production**

In our recent analyses of students’ laboratory conversations, we were struck by the extent to which the development of discursive competence mapped onto some aspects of an existing typology of sign production (Eco 1976). Here, we retain from Eco’s classification only those aspects that are relevant to the analysis of communication from our database and we modify individual categories to make them consistent with the evidence in our database. The remaining categories are mapped in Figure 5. We begin by introducing each category and then proceed to articulate the changes in modes of sign production and modalities of expression as students become increasingly familiar with some perceptual field (e.g., produced during investigation).

Eco (1976) classifies the modes of sign production according to the physical labor required to produce an expression ranging from simple recognition, to ostension, replica, and invention of previously non-existent expressions. ‘Recognition occurs when a given
object or event, produced by nature or human action (intentionally or unintentionally), and existing in a world of facts as a fact among facts, comes to be viewed… as the expression of a given content’ (Eco 1976: 221). We attribute to this category those instances in our database where students attempt to produce some phenomenon, simply looking at it (often without words), and sometimes commenting (i.e. ‘it doesn’t work’, ‘we can’t get it to work’, etc.). In this case, students are simply concerned with producing the phenomenon rather than developing theory.²

Eco (1976) notes that ‘Ostension occurs when a given object or event produced by nature or human action… is ‘picked up’ by someone and shown as the expression of the class of which it is a member’ (p. 225). He goes on to argue that ostension constitutes the most elementary form of active signification and uses the example of holding up one’s shoes not to say ‘shoes’, but rather ‘my shoes are dirty’ so that the object is both sign and referent. In this case, it is as if the speaker said ‘«shoes (ostension) + these (mention) + shoes (referent)»’ (p. 225). In our database, there are many instances when students, asked about what they have found out, simply utter something like ‘look, I show you’ and redo the entire investigation (silently, or simply naming some objects). Because the selected object (investigation) is selected as a whole to express its class, we categorize this as an ‘example’. When students show only part of the investigation to stand for the entire investigation (and gesture and describe omitted parts), we classify the situation as ‘sample’.
Replicas constitute a class of production modes that govern the most usual elements of sign production. Here, one takes account only of ‘replicable objects intentionally produced in order to signify’ (Eco 1976: 228). This definition covers all those expression units that use a continuum more or less alien to their possible referents and are correlated to content units in an arbitrary way. These expression units include (in addition to verbal devices), ideograms, coded kinesic features (gestures meaning «no», «yes», «that way», etc.), musical notes, traffic signs, symbols in logic and mathematics, and so on. In our database there are numerous instances where students use arbitrary objects to stand for the entities that they talk about. For example, a student might pick up three pens, then identify two as transparency films that are rubbed; she then identifies the third pen as standing for atoms that can be separated into nucleus and electrons (she pulls the body from the cap). Subsequently, the entities (films, atomic nucleus, and electrons) are no longer referred to in gesture or speech, but signified by means of the pens. Here, the objects used to signify have an arbitrary relation to their referents and therefore function in communication in the same way abstract words or gestures do. They are combinational units. When these objects are taken from the equipment of the investigation, they signify the objects and events to be explained in a stylized form. Finally, we categorize as ‘structured events’ those instances when students have arrived at a new way of perceiving and explaining some aspect of the world, which, in investigations, they often produce and see for the first time.

Evolutionary Trends in the Ontogenesis of Explanations

As students become increasingly familiar with some phenomenon, from initial recognition to the point of evolving an adequate and viable language for describing and
explaining the phenomenon, there is a change in the modality of expression (Fig. 4). (As indicated earlier, we are not so much concerned with the question whether the explanations are consistent with some external norm, but with the local consistency of the three modalities) Figure 5 shows that the emergence of scientific language begins with a (‘raw’) stimulus at the level of perception. Students begin by manipulating objects and watching events, frequently without commentary. When they are subsequently asked to account for what they observed, they often redo (‘We’ll show you’) the investigation and thereby provide an example (near copy) of the event.

As they continue in their attempts to account for observations and start to provide theoretical descriptions, students begin to enact parts of the investigation by means of gesture including possible observations (samples). In subsequent attempts, gestures are increasingly used against the equipment (no longer enacted manipulations) and a large number of deictic expressions (gestural, verbal) develop. As their inquiry continues, arbitrary objects (e.g., pens, rulers, and notebooks) are used as placeholders for objects involved in the experiment or conceptual objects (e.g., electrons).

Finally, students provide observational and theoretical descriptions with a minimum of gestures and in the absence of the experiment. In the ultimate stage (e.g., during exams), students provide complete verbal descriptions for the manipulations, events, and explanations. Running parallel to this development is a second developmental sequence in which students move from simply ‘naming’ objects (gestural and verbal deixis, words, and arbitrary objects) to constructing observational descriptions of specific events, to constructing observation categoricals and, finally, to providing complete explanations. In
this second sequence, gestures often precede the verbal modality in each of the four categories (naming, specific observation, observation categorical, and full explanation).

While reliance on the phenomenon steadily decreases (moves from the phenomenal towards abstract representations of the phenomenal) there is an increasing involvement of the gestural and verbal modalities. However, near the end of the sequence reliance upon gestures as an expressive modality also decreases; consequently, communication is comprised almost exclusively of the verbal modality. Students describe in words (and sometimes use gesture or diagram) the phenomenon and provide a scientific explanation (includes writing and diagrams) for it. In the following section, we provide detailed descriptions of case materials (from our database) that illustrate this emergent process of scientific language.

**From ‘Raw’ Stimulus to Scientific Language**

In this section, we provide detailed examples to portray the developmental patterns in the evolution of communication that we discerned in our database. In order to chart this evolution we adopt a ‘phenomenalistic epistemology’ (Quine 1995). Such an epistemology requires a capacity for recognizing perceptual similarities.

Following Quine, we hold that global stimuli are perceptually similar when they have a shared salience. Furthermore, salience is crucial for picking out objects from the background noise of the situation. Here, motion assists in this process, particularly forms of motion that can be produced in the presence of the interlocutors. ‘Sweeping or pointing gestures in the vicinity or direction of the intended portion of the scene thus implements the desired association with the spoken word’ (Quine 1995: 18-19). To arrive at socially shared perceptions two individuals have to witness two consecutive scenes and
experience them as perceptually similar, and communicate that they experienced them as such. Signals (verbal, gestural) are but part of the perceptual salience that gets correlated with other signals and with situations. Beginning with a form of talk that we term ‘muddle’, the trajectory then moves from observation sentences to observation categoricals, and to the simplest forms of explanation (theory) that generalize correlations across situations (‘Whenever it rains, there are clouds’).

‘Muddle’ and Conceptual Mayhem

Students do not come to school as complete blank slates. They bring with them perceptual, gestural, and verbal competencies evolved from and through their participation in the culture of which they are part. These competencies serve as resources in the novel situations faced by students in school science laboratories. However, these resources are often inappropriate for seeing and understanding phenomena as scientific, and often impede the development of scientific discourse—as the research on ‘misconceptions’, ‘alternative conceptions’, and naïve theories has shown (Pfundt and Duit 1994).

In our studies, we also observed a proliferation of linguistic resources that students bring to the learning situation and use either inconsistently or inappropriately (though they do not know this). For example, there was considerable variation in the words and phrases used for the different elements in the microworld and in understanding how these elements (object, arrows) interact. The result is a form of talk which we refer to as ‘muddled talk’ (e.g., Rorty 1989). This is evident in the following brief episode involving the three students E, G, and R who investigate a Newtonian microworld. The three students look at a computer screen display not unlike that featured in the appendix. The
task pursued by the three students is to describe and explain the events displayed on the
monitor.\footnote{1}

01 R: The big one is time step.
02 E: Which one is time?
03 R: It’s that big arrow.
04 G: Oh yeah, the big arrow is time. "The big operator is time."
05 E: So then the little arrow is direction?
06 G: Yeah, big arrow is direction. No, I mean, big arrow is velocity.
07 E: Is time.
08 R: No, it’s time, but it also directs, though.

Here, there exists confusion about the nature of the two arrows including how to refer
to them. Unbeknownst to the students at this time is also the fact that they use ‘little
arrow’ or ‘big arrow’ to denote different objects. For example, R uses ‘big arrow’ to
denote «force», where ‘big’ is associated with the volume of the arrow. G, on the other
hand, uses ‘big’ in the sense of tall or long, and therefore denotes the longer arrow as
sometimes «force» and at other times «velocity». It is not surprising then that students
take considerable time even to refer in consistent ways to name the objects that are
salient. For example, one group of students referred to «force» using eight different
lexical items {‘little arrow’, ‘big arrow’, ‘time-set’, ‘time’, ‘direction’, ‘time-and-
Similarly, they referred to the other arrow, «velocity» using eleven different terms {‘little
arrow’, ‘big arrow’, ‘initial speed’, ‘velocity’, ‘force’, ‘effort’, ‘strength’, ‘speed’, ‘direction’, ‘speed-and-direction’, and ‘velocity’}. What complicates the situation is that students often do not notice when their peers use different lexical items to denote the same perceptual objects, or use the same lexical items to denote different objects. Science classrooms are not unlike present-day versions of Babel. In this seeming chaos, deictic and iconic gestures were crucial to establishing a common ground, perceptually isolating relevant features, finding appropriate descriptions, and ultimately arriving at a viable theoretical language.

Picking out Entities (Salience, Gesture, Motion)

A crucial element in the evolution of a common language about scientific phenomena is the establishment of shared perceptual elements, the objects and events that the observational and theoretical language are intended to describe. The greatest challenge for the students is to structure the perceptual field and to come to some agreement that their respective perceptual structuring can be taken as the same, for the practical purpose of their collective work. In this situation, pointing to and gesturing provides an important resource to align a coparticipant to something one has identified even if a descriptive language for navigating the world does not yet exist. Pointing to and gesturing are crucial to enhancing salience, picking out and communicating entities from the scene that the talk is about. Therefore, it is not surprising to see students, especially in the early stages of their investigations, frequently use the gestural modality in their communications. In this, the episode in Figure 6 is typical, showing the hands of three students in front of the screen, each taking its part in the conversation (even if the third person does not talk). Here, J (to the right of partially visible center person) gesturally enacts ‘straight up’
(Frames 1 and 2). The latter part of his (gestural and verbal) proposition is overlaid by
N’s (to the left of center person) contribution (who is vying for the turn), which also
consists of verbal and gestural contributions. Giving up his turn, J both stops talking and
removes his index finger, which was already in position for a subsequent proposition.
Here, we can see turn taking enacted not only in the verbal modality (a much studied
phenomenon in the domain of conversation analysis) but also in the gestural modality.

Gestures are, at this stage of learning science, not just ancillary and providing
redundant information—as claimed by Crowder (1996)—but they are central to
constructing a semantic model and to successful communication. That is, particularly in
this situation where students are unfamiliar with the domain ontology and a language that
describes it, the working together of the three modalities discussed here is crucial. For
example, in the following episode (Figure 7), G asks his partner R to modify the
eperiment in a particular way. Without the perceptual and gestural resources, R would
be hard pressed to know what G is talking about. However, when the three modalities are
read together, we can understand his utterances as a request to attach «force» (which R
currently ‘holds’ by means of the cursor) and orient it in vertical direction upward and
then turn «velocity» so that it is also oriented upward.
In the first instance, G moves his index from touching the ball upward and parallel to the screen. Although G has not specified in this episode what ‘it’ is that is to be directed upward, the situation is made less ambiguous to a certain extent. «Force» has been the object that the conversation is about and R currently holds it with the cursor. G begins the second part of his utterance with a ‘and then’, which suggests a second request, this one concerning the second object clearly identified here by the gesture. Although G has difficulties expressing himself in words, his gesture (Line 02) enacts what one has to do with the cursor to move «velocity» into an upward position: his index touches the tip of the arrow, then the hand turns until the index points upward. Here, the gesture simulates an action that is not communicated in the verbal modality: the hand ‘rotates’ «velocity» whereas the utterance simply suggests to ‘put… upward’. Finally, in Line 03, the entire hand moves upward with the index continuing to point into the same direction. Because the index already showed the direction in which «velocity» was to be turned, Line 03 can then be read as a gesture that projects what will happen if the experiment is set up the way Lines 01 and 02 suggest: the circular object will move vertically upward.

In this situation, the language is insufficient to communicate on its own. Even with visual access to the scene it would be difficult to understand. However, together with deictic and iconic gestures, the three communicative modalities taken together provide students with important resources for bootstrapping the process that will lead them to scientific language. There is a further interesting aspect in this episode, a phenomenon that we have previously reported relating verbal and gestural modalities (e.g., Roth 1999b, in press). During learning phases, one can observe a delay between the verbal and gestural modality such that the latter precedes the former. Here, G suggested to ‘put…
upward ‘velocity’. Although McNeill (1992) suggests that the agreement between verb and gesture is to be analyzed, ‘put’ in itself does not specify what to do with the arrow. Rather, viewing the episode on video gives the impression that the ‘upwards’ is delayed in forthcoming. We therefore view the gesture, intimating a turning of ‘velocity’ in the upward position as being completed considerably before the utterance that provides a meaningful description of the action.

When students do not have a common language for picking out entities or it is not certain they are attending the same entities, pointing (deictic reference in the gestural modality) becomes a central resource in coordinating the three modalities. This is exemplified in the next episode (Figure 8) which represents one exchange between the two students J and M. A third student, B, has the mouse with which the on-screen entities are modified. J suggests that B has to click on the tip of the ‘big arrow’ (achieved by grabbing the ‘handle’ which allows an arrow to be changed in magnitude and direction) to change ‘its direction’. But from J’s suggestion it is not clear exactly what he wants to be done. The longer ‘velocity’ arrow can be changed (in magnitude and direction) by ‘grabbing’ the black dot at its tip. The shorter but wider ‘force’ can be changed in magnitude and direction by grabbing and moving the black dot at either end. Furthermore, ‘its direction’ may actually refer to the direction of the circular object, which may be changed by changing one or the other arrow. M’s response reveals what he had heard J saying.

]]]]]]]]]) Insert Figure 8 about here [][[]][[][]]
M responded that to change the direction, B had to click ‘here’, while he simultaneously touches the screen right at the back end of «force». This deictic gesture constitutes «force» as the signifying figure against a more diffuse ground, and more specifically, the black dot at the end of «force». He therefore indicates that a change in the direction of «force» is accomplished by manipulating the arrow with the handle salient as a black dot. M therefore does not just point out a particular spot. Rather, the black dot motivates the utterance ‘here’ and the pointing gesture as much as the verbal and gestural deixis motivates the perceptual salience of the black dot. In essence, then, M hears J talk about «force» and, in providing a different description of the action for changing the direction, both corrects J and displays what he understands the ‘big’ arrow to be by unambiguously pointing to one of the two arrows («force»). This episode is therefore an example how people deal with the ambiguity of deixis in praxis. Given that there are two arrows in this situation, there is a potential for alternate referents. M therefore engages in a conversational repair because it is not clear which object is being referred to and where the action is to take place.

Observation Sentences

Students’ first viable observation sentences are highly context sensitive in the sense that interlocutors and analysts have to have access to all three modalities to understand what is being communicated. The next episode (Figure 9) exemplifies the nature of students’ observation sentences.
At this point, different students still use the same signs to signify different entities. Yet the gesture and pen limit the interpretive flexibility of the communication as a whole. The pen motivates our noticing of the ‘little arrow’, and is at the same time motivated itself by the arrow. In the first five panels, G holds the pen parallel to the arrows. As the hand moves from left to right, the pen ‘always stays straight’. In the second part of the utterance, G then appears to talk about the circular object which ‘goes in that straight direction’. Here, somewhat confusing the matters, the hand moves through a curvilinear trajectory and away from the screen, the pen no longer in an orientation that ‘always stays straight’. The pen can be interpreted as another interpretant of the ‘little arrow’. Although in this case the pen is a motivated interpretant, in contrast to other episodes in our database where pens and other objects had an arbitrary relationship to their referents (e.g., ‘electrons’ [Roth 1999a]).

Reading across the three modalities, we can understand this episode as consisting of two observation sentences. The first observation concerns one of the two arrows that ‘always stays straight’. Here, the present perceptual modality is inconsistent with the verbal, for it only shows a single state rather than a series of states that would motivate the descriptor ‘always’. The orientation of the pen, which stays constant across the three frames of the utterance, recalls an observation made immediately prior. In the second part of the utterance, G makes a verbal assertion about something (‘it’) going in a straight direction, which can be heard as an assertion about the object. At the same time, the gesture seems to be inconsistent with this reading, for frames 6-8 show that pen and hand do not remain parallel to «force» and trajectory. However, given that G sits to the left of
the computer, his arm is actually extended to its maximum so that any motion would be curvilinear unless he bends forward or gets out of his seat.

Observation Categoricals

As students continue to explore the focal situations, they begin to formulate relationships, particularly causal relationships. Here again, these formulations do not appear at first in the verbal modality, but as an assemblage of the three modalities. The next episode (Figure 10) exemplifies the nature of initial observation categoricals.

In this episode, G’s gestures already provide descriptions and explanations, which are more appropriate (and therefore more viable in the ecology of the school context) than his verbal descriptions. Here then, his gestures considerably preceded his verbal representation of Newtonian physics. In this situation, gestural (and verbal) deixis was crucial in coordinating utterances, gestures, and the phenomena in the microworld. At the time of this episode, G and his classmates had not evolved a language to consistently describe the circular object and arrows in internally consistent (and even less in scientific) terms. That is, although the words ‘force’ and ‘velocity’ may have appeared in their conversations, they were not used in a manner consistent with the scientific norm. (They used the appropriate scientific [verbal] discourse only two weeks later during a subsequent lesson with the microworld.) When we look at the gestural modality in the present episode, however, we can interpret the relations as correctly describing the relationship between the velocity and force as the corresponding arrows («velocity» and
«force») represented them. In Figure 10, we see that the fingers of the right hand parallel to «force» (frames 2-5) always stay parallel. In contrast, the fingers of the left hand (frames 3-5) which are initially parallel to «velocity», change their direction in the way «velocity» could be seen to change on an earlier screen during the immediately prior experiment. That is, the hands, as signs standing in a reflexive motivating relationship with the perceptual signs, describe the situation in scientific terms, although the (verbal) language is yet inappropriate.

Nevertheless, the verbal modality already uses everyday terms commensurable with the (scientific) language these students later develop. That is, although G does not yet describe the events in terms of force, acceleration, and velocity in ways commensurable with Newtonian physics, he characterized the action of «force» as ‘pushing’, which is a vernacular form of describing forces. He also associated the longer «force» with a resulting higher velocity. Here, the referent of ‘velocity’ is not entirely clear and two hearings are possible. Because the utterance coincided with the positioning of the left hand, ‘velocity’ can be heard as an interpretant of the left hand: therefore, the longer «force» pushes more and therefore leads to a longer «velocity». But the fragment ‘Since that arrow’s longer the velocity is higher’ can also be heard as an interpretant for the longer «force» as equivalent to a higher velocity. In this case, ‘velocity’ would be his interpretant of «force» (incorrectly so from a scientific perspective). However, the nature of the interpretants for each of the two hands is clear, if their position in space in the course of the motion is seen as bearing a motivated (iconic) interpretant relationship to the signs in the perceptual modality. If seen in its entirety across the three modalities, G’s
description and explanation of the events was scientifically correct long before he was competent enough to construct explanations in the verbal modality alone.

In this episode, the other students are able to make sense of G’s utterance although he used the same indexical terms in different ways. First, ‘that’ appears three times and ‘it’ appears twice. In the first instance (frame 2), ‘that’ has a deictic function making salient and linking «force» and fingers of the right hand. In the second instance, ‘that’ (frame 5) introduces the causal consequence (‘that’s why’) of the hand arrangement he had set up. This causal relationship constitutes the observation categorical and is part of an argument. Finally, in the third instance, ‘that’ coincides with the double gesture of the moving and changing arrows against the background of the screen. Here, ‘that’a way foregrounds the curvilinear trajectory presently available only in the gestural modality.

Second, G twice employs the indexical locution ‘it’. Although this might lead to confusion, the coincidence with gesture disambiguated the two differing referents (‘It’s pushing it…’). Because the right hand perceptually follows the left hand, the former is literally pushing the latter (frames 3-6).

Our analysis and the events immediately following this episode also highlight another aspect related to gesturing. Although analysts (and in practical situations teachers, psychologists, etc.) may attribute a particular meaning to communication, some communicative acts may not be understood in practice by the interlocutors. Here, neither R nor E appear to understand the relationship between the arrows on the screen and those in G’s utterance. Consequently, both engage G in a conversational repair, which itself draws on multiple modalities.

R: What?
E: Which arrow is longer?

G: Do you know how we had this one bigger than that one.

Points to «force» Points to «velocity»

R’s utterance ‘What?’ is simply a generic indicator for not having understood.

Interestingly enough, although one might assume the mutual availability of the objects in the perceptual modality, E asks which of the two arrows is longer. In response, G uses verbal and gestural deixis to isolate the two objects involved in the comparison ‘this one bigger than that one’.

Two weeks later, amounting to two additional lessons of experimenting with the microworld, many students had arrived at making observational and theoretical (observation categorical) descriptions entirely in the verbal modality. For example, as they prepared for another task in which they were asked to design demonstrations that would teach grade 8 students about motion, G suggested:

We’ll design the experiment based on Newton’s law… the opposing forces, one has two equal but opposing forces, and then there is no change in velocity. And if there are unequal and opposite forces, objects will accelerate in the direction of the greater force.

In this excerpt, G provides a scientific observational and theoretical description of two situations. If there are two equal (magnitude) but opposing forces, there is no change in velocity. However, if there are unequal and opposite forces, there will be an acceleration in the direction of the greater (in magnitude) force. At this point, students no
longer need the perceptual and gestural modality to construct their observational and theoretical descriptions. The verbal modality has entirely taken over the task of sign (sign complex, text) about some event in the (Newtonian micro-) world—the relationship between opposing forces acting on an object, on the one hand, and the velocity and acceleration of this object, on the other.

**Discussion**

**Multi-Modality of Communication**

Traditionally, domains such as semiotics and linguistics have focused predominantly on language as the primary modeling system, often considering non-verbal means as secondary, derivative, or partial translations of the primary system. In contrast, our videotapes show that to beginners, signs in all three modalities (including verbal modality) are attended even though they are highly underdetermined and variable. Thus, students do not structure the computer or experimental displays in the way a physicist or physics teacher might do. Furthermore, students also differ among each other in terms of what they determine to be salient. For example, despite the relatively simple configuration in the figure, it was not notable to students at first that the outline arrow («force») did not change in magnitude or direction, that the line-arrow changed both magnitude and direction, or that the trajectory of the object was parabolic. The figure-ground ensemble has to be taken as variable, because in the perceptual processes of the individuals, they are variable. Our data also show that the verbal signs are variable. Students used up to 14 different words and word combinations to refer to each of the arrows before they settled on a particular one. Furthermore, some of these (e.g., little
arrow, big arrow, direction) were used to denote both «velocity» and «force». It is in the context of particular utterances, or the background that the sign takes shape in the same way that some aspect of the ground takes shape and becomes figure. For example, if there is an arrow pointing upward behind the hand, the sign-function relates the direction of the finger and that of the arrow. (In a different context, the same configuration may indicate testing for wind striking the finger horizontally.)

Observation and theory expressions made concurrently across all the three modalities are more complex than in individual modalities. However, over time, and with increasing familiarity, the verbal modality assumes the function of primary representational modality. (At that point, students can write about some scientific investigation even though they are absent from the setting and materials where they conducted it.) We moved to the consideration of the three modalities (verbal, gestural and perceptual) motivated by Eco’s (1984) description of signing as a process in which some ‘lumps’ from the dynamic continuum come to stand for other ‘lumps’. As a result, the relationship between the different lumps is symmetrical: ink on paper in the form of ‘tree’, a sound (generated by human or computer) of ‘tree’, a line drawing of a tree, or a tree in the garden each may stand for another during communication. Therefore, we can think of students’ communication as running in any one or more of the three modalities. There is the world before them from which particular entities are made salient by means of utterances and different types of gestures. There are the gestures that tell part of the story, and there are the words. The difficulty arises in part from the fact that the gestures and words can be shifted during periods of transition (that is, learning in a new domain). What our videotapes reveal is a multi-level process. First, there are the processes of
shaping the content and expression (gesture, language) continua. Second, there are the processes by means of which expression continua are correlated with their possible content. Finally, there are the processes by means of which signs get connected to the segmentations of the content continuum.

Non-verbal signs have to be considered in a theory of communication because of their prevalence in face-to-face conversation (Bavelas and Chovil in press) and particularly in those interactions when the conversational topic is a feature within the setting itself. Certainly, language is the most powerful semiotic device (Eco 1976), but there are semantic spaces that it does not cover as effectively as other devices (Lemke 1998). In other words, the indexical ground is intimately tied to basic processes of human interaction and participant frameworks. We show here that leaving it out of the analysis of communication means omitting one of the resources of semiotic processes. This observation parallels Hanks’ (1992) suggestion that the indexical ground shapes the interaction, which in turn shapes the indexical ground. However, in our research we are less concerned with the interaction per se and more with the nature of communication, particularly the production (and making salient) of signs in communicative processes.

Both gestural and verbal modalities have a deictic feature that can be used to foreground the perceptual modality within the interaction. Some readers may assume gestures to be somehow unambiguous, and therefore sufficient to ground some utterance. They are indexes or signs that ground utterances, or expressive media. But this is not so. The shape and direction of the pointing is itself ambiguous so that what is being communicated arises from the interplay of percept, utterance, and gesture. Each constraints the flexibility of the other through the co-occurrence of other modalities.
From this triangulation of potentially different meanings, a lower number of specific meanings arise.

**Emergence of Scientific Language**

Students such as those in the present episodes are in a fundamental dilemma in that they do not know the domain and have no (linguistic, perceptual) dispositions for structuring the field in front of them. What can be conceived from a particular perceptual field is fundamentally problematic and not determined by the nature of the material of the field; rather, the nature of the entities arises from cultural structuring practices (e.g., Goodwin 1994; Livingston 1999). Thus, when students conduct new and unfamiliar investigations in science laboratories, we have the paradoxical situation that they have to establish expressions for a content model that does not yet exist as such. As such, our students are engaged not simply in learning a language about a given world but are fundamentally engaged in an emergent process of developing scientific explanations. For this particular group of students this process involves developing a set of linguistic and perceptual structures that are consistent and coherent across a variety of different situations that can be created with and in the microworld.

In this article we provide examples of the learning trajectories of students in science laboratories that lead to the emergence of scientific language. Students begin their learning trajectories with what can best be described as ‘muddled talk’ and end up, given as much time as in these lessons, with viable ways of talking (and writing) about phenomena in observational and theoretical terms. That is, from the almost chaotic and incomprehensible utterances emerges a viable language. (Although the role of the teacher
and other scientific resources is significant in this process [e.g., Roth 1995], space
limitations prohibit an extension of our analysis to that level.)

Once students have developed consistent ways of verbally representing particular
entities, the use of gestures decreases. For example, when they begin to call «force»
‘gravity’, they no longer need to use the ambiguous ‘arrow’ with the accompanying
dectic gesture. A simple command such as ‘no, no, no, pull on the gravity’ suffices to
make salient and direct the person’s attention to the appropriate arrow. If, as we propose
here, students’ communication is analyzed in terms of three modalities (perceptual,
gestural, and verbal), we can observe a shift from communicative efforts initially based
on the perceptual modality, moving to heavily mixed, and ending in verbal modalities. In
the process, verbal and gestural deixis links the three modalities, sometimes providing
seemingly redundant information, at other times seemingly contradictory information.

Our analysis makes it explicit that we do not conceive of learning as the ‘taking in’ of
information. Rather, we have portrayed learning in this article as the coordination of
signs that appear across modalities. Our project is decidedly phenomenological in that we
are interested in understanding the initial conditions required for the formation of cultural
units. Because communication requires such units, these cultural units are accepted as
data in traditional semiotics. For our research participants, however, this assumption
cannot be made. They are at a point where field and language, referents and signs,
constitute extremely ambiguous rather than explicit messages. Cognitive development
and learning are therefore constituted in the evolution of a structured and coherent set of
perceptual field and expressive means.
From our research it appears that the gestural modality is of particular importance in the process of emergence of scientific language. Gestures, involving the human body, are of the same type as the events that students are asked to describe and explain. Therefore, gestures and perceptions, viewed as signs, motivate each other assisting in making salient objects and events so that they, initially intra-subjective are made accessible on an inter-subjective level. Our research shows that students appear to learn to talk science more readily when they are provided with opportunities for communicating in the presence of material experiments (Roth and Welzel in press) or visual representations (Roth 1996). The theoretical argument presented here would account for this facilitation. In the presence of the materials and diagrams, there are simply more communicative modalities. Furthermore, these other modalities are co-constitutive of the entities re-represented in the verbal signs.

Iconicity

Research on the nature of iconic reference and the level of iconicity is of considerable practical interest in at least two domains. First, developmental researchers who attempt to infer children’s understandings from their utterances and gestures have to take account of Goldin-Meadow et al’s (1992; 1993) research that showed that during change in understanding there is a discrepancy between gesture and utterance. To make appropriate inferences about what a gesture means we need to know the relationship between the gesture as sign and what is being referred to. We suggest that this needs to be done from the perspective of the learner who does not yet know the ultimate semantic model she will construct. Second, teachers who want to assist students need to make appropriate inferences from the gesture to scaffold the emergence of the verbal expressions.
associated with some concept being gestured. In these instances, individuals are said to rely on ‘iconic’ signs to read what their interlocutors do not and cannot make available in the verbal modality. This puts enormous emphasis on the interpretation of iconicity. To extent our discussion of iconicity and its problems, we elaborate the following examples.

The gestures such as those in Figure 8 appear unambiguous and non-problematic. The index pointing upward (Line 01), an image of an arrow aligned in this direction, the ‘turning’ of «velocity» (Line 02), and the moving hand (with pointing index) predicting a particular motion (Line 03) are all forms of gestures that can be categorized as ‘iconic’. Particularly when there are possibilities to compare the gesture with the signs in the perceptual modality occurring immediately before or after. In the episodes depicted in Figures 9 and 11, however, categorizing the gestures as ‘iconic’ is much more problematic. In Figure 9, we saw that G’s utterance ‘It goes in that straight direction’ was accompanied by a gesture in which the hand holding a pen moved from left to right across the screen (frames 6-8). However, the pen did not remain parallel to «force» as previously suggested and the trajectory of the hand (interpretant of perceptual ‘object’) did not move straight—as the object had done earlier—or parallel to the earlier displayed trajectory.

[Insert Figure 11 about here]

On the surface, the episode constitutes an observation sentence about some arrow that ‘just goes’. Here, given that the previous experiment involved only «velocity» we can understand it as the interpretant of ‘arrow’. Furthermore, the previous image on the
screen showed—to the researcher—constant «velocity» (magnitude and direction) in which case ‘just goes’ could be read as an interpretant of ‘unchangingly’. However, the gesture does not seem to bear any relation to the events on screen. The hand moves such that the arm rotates around the elbow (Line 01) and, in the end, descends (Line 02). The hand, therefore, described a curvilinear trajectory away from the screen. As such, it is not ‘iconic’ to anything in the perceptual modality that could have served as the motivating counterpart of the gesture. The gesture evokes a curvilinear trajectory while the verbal modality provides the description ‘the arrow just goes’.

Interlocutors likely operate under the assumption that the three modalities are consistent with each other. The interpretive flexibility associated with the signs in each modality is therefore constrained. On the other hand, in periods of transition the different modalities can express different entities and processes (Goldin-Meadow et al. 1992; 1993). At these moments, the nature of the entities and processes in the world may be different; what is salient is a matter of the historic constitution of individual perceptual processes. In many cases, selection has evolutionary precedents; however, this is not to suggest that perception itself is a fixed process.

Our research also provides some indication that the notion of iconic gesture is in some ways profoundly problematic (cf. Eco 1976; Fenk 1997). These problems result in part from the fact that ‘iconic signs’ are only partially ruled by motivation (similitude); the other part results from conventions and stylistic rules. Some phenomena initially referred to as iconic often turn out to be something else. Furthermore, iconic signs cannot be classified as a unique category since they circumscribe other kinds of signs (subject to their own unique procedures) which in turn circumscribe iconic signs. The relationship is
transactional. Summarizing his extensive critique, Eco (1976: 216) concludes, ‘iconism is not a single phenomenon, nor a uniquely semiotic one’ and suggests replacing the typology of signs (iconic, indexical, and symbolic) by a much clearer typology of modes of sign producing functions.

Despite the incisive critique of iconicity and the nature of the iconic sign, research concerned with iconic gestures seems to be untouched by the problematic nature of iconicity. A fundamental question that arises from our research therefore concerns the nature and degree of ‘iconic’ in iconic gestures. An iconic gesture has been defined as ‘a formed gesture which depicts in its form or manner of execution aspects of the event or situation being described verbally’ (McNeill and Levy 1982: 275). These issues give rise to new questions including: ‘How closely must the gestures be aligned so that they can be treated as iconic to the thing it portrays?’, ‘Is this a practical problem solved by the interlocutors?’, ‘What are the forms permissible in the gesture so that they can be characterized as iconic to the entities that they are said to be about?’, ‘How do we interpret gestures when, as Goldin-Meadows and her associates suggest, there is a disjunction between verbal and gestural modality?’, and ‘What is the role of gestures in interaction and how interlocutors deal with the interpretive flexibility inherent in gestures?’ These and other questions remain to be answered by detailed studies on the emergence and evolution of structured perception and communicative means.
Appendix

Physics of Motion and Computer-Based Newtonian Microworlds

The students in the episodes were taking an introductory qualitative physics course where they also were provided with the opportunities to use a Newtonian microworld. This microworld presents images in which physical phenomena appear in the same plane as the conceptual entities that explain their motion (Figure A.1).

In a Newtonian world, the motion of an object is fully determined when its initial position and velocity, and the forces acting on it are given. The following equations relate position $x$ and velocity $v$ to the initial position $x_0$, initial velocity $v_0$, force $F$, and mass $m$ of the object. (Note that position, velocity, and force are vectors, a fact that is signified by bold-facing the corresponding letters.)

$$ x = x_0 + v_0 t + \frac{F}{2m} t^2 $$  \hspace{1cm} (1)

$$ v = v_0 + \frac{F}{m} t $$  \hspace{1cm} (2)

Students who learn physics have difficulties learning the relationship between the relevant quantities. Often, they learn to manipulate the equations without having an understanding of the underlying conceptual entities and their relations. Interactive Physics™ is a piece of computer software that allows students to explore the motion of phenomena objects in the copresence of conceptual entities. Vector quantities can be represented using arrows, which have a certain degree of similitude with the way we experience the world. Thus, outline and single-line arrows signify force and velocity, respectively; direction and length of an arrow signifying direction and magnitude of the
referent entity. The screen print in Figure A.1 shows the interface students use, and a simple experiment. a. Top left. A circular object with an initial velocity pointing nearly upward. A force pointing to the right is attached to the object. After students push the ‘start’ button (top left of the screen), the object is set in motion according to the underlying Newtonian principles given the chosen velocity, force, and mass of object. The figure shows that the trajectory of the object is parabolic, a fact also represented in the quadratic form of (1).
Notes

Grant 410-99-0021 from the Social Studies and Humanities Research Council of Canada and a SSHRC internal grant from the University of Victoria supported this work.

1 To avoid the confusion that reigned initially among students (see section ‘Muddle and Conceptual Mayhem’), we use the notion «velocity» and «force» to designate those arrows (vectors) that scientists would recognize as velocity and force, respectively.

2 Elsewhere, we described the fundamental double bind that students find themselves in when they produce phenomena following some set of procedures (Roth et al. 1997a). First, in order to know that what they have done is what they were supposed to do, students need to know that what they have seen is what they are supposed to have seen. But, second, in order to know that what they have seen is what they were supposed to see, students need to know that what they have done is what they were supposed to do.

3 We use the following transcription conventions:

Yeah: Italicized words were stressed in the utterance;

°big operator°: Degree signs enclose low volume, almost inaudible speech;

[«force»]): Double square parentheses enclose the object indexed by the speaker;

..?: Punctuation marks reflect how utterances are heard (completed, as questions) not grammatical units.
References


Captions

Figure 1. Different models of sign production. a. Butterworth and Hadar suggest that gestures aid in the lexical search (itself driven by the semantic model), but constitute a different channel. The four processes labeled with Roman numerals are discussed in the text. b. In the McNeill (Bates, Eco) perspective, an underlying semantic model drives the generation of arbitrary expressive units (utterances) and transformations by similitude (gestures). There are suggestions that gesture precede and are the basis of language.

Figure 2. Our model constitutes a modification of that proposed by Eco (1976) for sign production during invention phases. It takes account of sign production in situations where individuals do not have a semantic model, and where perceptual continuum and expressive media are unshaped. Through manipulation of the continuum (experimentation), perceptual model and expressions (gestures, language) are mutually adjusted ultimately resulting in a semantic model.

Figure 3. Our unit of analysis includes three concurrent modalities during communication. The interaction between hand-finger in the pointer configuration and the perceptual ground (left frame) leads to the salience of the vertical arrow (velocity).

Figure 4. Changes in modes and modality in which signs are produced as students become increasingly familiar with some perceptual field from which they learn to perceptually isolate some phenomenon that they learn to explain. The number of squares is a qualitative indicator of the extent to which a particular cell is prevalent during communication.

Figure 5.

Figure 6.
Figure 7.

Figure 8.

Figure 9.

Figure 10.

Figure A.1. Interface of the Newtonian microworld used by the students in the episodes. The outline and single-line arrows signify force acting on and velocity of the circular object. Top left. Initial conditions that obtain before the investigation. Bottom right. Flash photograph capability allows students to freeze the positions held by the object over time.
a.

expression units

semantic model

b.

stimuli

perceptual model

perception

abstraction

similitude

transformation

(i.e., gesture)

arbitrary

independent set of expressive units

(literatures)
M: See that is not the force,

: This one is the direction

(J handles «velocity», the vertical arrow.)
perceptual

gestural

'this' 'this'
is the direction'
<table>
<thead>
<tr>
<th>Modality of Expression</th>
<th>Modes of Sign Production</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recognition</td>
</tr>
<tr>
<td></td>
<td>Unstructured Events</td>
</tr>
<tr>
<td>Percept</td>
<td></td>
</tr>
<tr>
<td></td>
<td>■■■■■■■■■■■■</td>
</tr>
<tr>
<td>Gesture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>■■■■■■■■■■■■</td>
</tr>
<tr>
<td>Speech</td>
<td></td>
</tr>
<tr>
<td></td>
<td>■■■■■■■■■■■■</td>
</tr>
</tbody>
</table>

Time (Increasing Familiarity)
J: Put the gravity straight up
N: The gravity is going up now
Try to put it upwards, and then put tha, tha upwards.
J: If you wanna change its direction, click on the tip of the big arrow.

M: No, if you want to change the direction, you click here.
The little arrow always stays straight. Whenever we get it

Like when we are doing it. It goes in that straight direction
Wouldn’t the length of the arrows (1.6) Since *that* arrow’s longer the velocity is higher. *That’s* why: it’s pushing it that’a way
G: The, the arrow just

G: Goes