Gesture and beyond

Nathaniel Smith
njs@pobox.com

Undergraduate Honors Thesis
Program in Cognitive Science
University of California at Berkeley

31 December 2003

1 Introduction

In recent years a great deal of interest has arisen in co-speech gesture, the meaning-carrying bodily motions that accompany ordinary linguistic communication. This research contradicts the traditional assumption that gesture is irrelevant to linguistics, arguing instead that speech and gesture production is tightly integrated, and that speech and gesture should be studied together as an interconnected system. Co-speech gesture is universal and ubiquitous across cultures, develops alongside speech (Butcher and Goldin-Meadow 2000), co-times with speech with millisecond precision (McNeill 1992), maintains this co-timing through disfluency, stuttering (Mayberry and Jaques 2000), even aphasia, and can serve many of the same functions as speech, including reference, discourse negotiation, and the expression of propositional content.

However, while the existence of a connection is difficult to deny, the details are as yet unclear, as are the implications for theories of language and mind. In this thesis, I revisit old questions about the relations between linguistic communication and the rest of humanity's cognitive apparatus, and argue that the barrier between language and thought is more porous than has previously been realized.

It is well understood by cognitive linguists that, contrary to generativist orthodoxy, language is not a discrete module entirely distinct from the rest of cognition; such a position is biologically implausible and contradicted by some decades of research on embodiment, grounding, etc. There is no corresponding cognitivist orthodoxy on the relation between language and thought, but a reasonable theory might hold as follows: semantics is embodied, and language is produced by brains and bodies; it is impossible in principle

to draw a strict line around language processing. Yet, there are things about speech that suggest it is special: aphasias can affect particular aspects of speech production in systematic ways and there is a critical learning period for language; language is universal and there are universalities to the structure of different languages; there are clear adaptations for language in humans both at the level of gross anatomy (e.g., the shape of the throat and jaw) and neural architecture (e.g., categorical perception of phonemes). Furthermore, speech production and comprehension has to happen extremely efficiently with strong real-time constraints; altogether, we might postulate, speech and gesture form an integrated system that in online processing works via specialized neural circuitry evolved for this purpose, and thus has only limited interaction with other aspects of cognition.

This is a sensible theory, but I will argue that it is wrong, by presenting data demonstrating that speech and gesture can interact in fine-grained ways with other *ad hoc* communicative mechanisms, aspects of the environment, and the conceptual system generally.

These claims are not without precedent. The existence of signed languages is already proof of a certain flexibility in the human language faculty, and among traditional speakers of the Australian aboriginal language Arrernte, speech is ordinarily accompanied by not just gesture but also drawing in sand, which systems are used together in a complex and partially conventionalized way (Wilkins 1997). Furthermore, a number of researchers have discussed the essential grounding of interaction in the environment (Goodwin 2000; Ochs, Jacoby, and Gonzales 1994; Ochs, Gonzales, and Jacoby 1996; Hutchins and Palen 1997; Clark 2003), and especially the complicated conceptual structure necessary to interpret deictic gestures (Hanks 1990; Haviland 1993; Hutchins and Palen 1997; Hindmarsh

and Heath 2000; LeBaron and Streeck 2000; Goodwin 2003; Clark 2003).

Clark (1996) argues that language use should be considered not as a phenomenon unique to itself, but rather as a particular form of "joint action". Roughly, the view is of language as providing a set of affordances for action and communication. Humans should generally not be understood to be engaged in speaking *per se*, but rather to be attempting to achieve some (possibly communicative) purpose, drawing on whatever resources (including linguistic) they have to achieve it, in interaction with others performing similar activities.

The present work follows in these traditions, but expands on previous results in several ways. Previous researchers have observed that the use of gesture relies on social context and the structure of the physical environs. In §3, I reiterate this point with new examples, and extend on it to argue that speakers not only use the environment for communicative purposes, but actively modify it in conversation to better support their gestures; that this modification is tightly integrated with speech production; and that supposedly extralinguistic phenomena can interact in a fine-grained way with such central pieces of online language use as backchanneling, turn taking, pause protocols, speech rhythm and timing, and the general structuring of discourse interaction. In §4 I expand further to give a definition and preliminary analysis of a particular phenomenon, "persistent structure", that shows systematic variation across modalities, indicating a general process at work. §5 explores persistent structure's uses in discourse, with some more detailed analyses in §6. Then in §7 I show how persistent structure is also used via offloading to assist in general cognition, and conclude in §8.

2 Data

For examples, I will draw on a new corpus of recordings of math students at UC Berkeley totaling approximately 20 hours to date. The students range from third-year undergraduate math majors to first-year graduate students, and were filmed working on a diverse collection of topics ranging from topology to logic and model theory to abstract algebra. A goal during collection was to make sure the data was as naturalistic as possible; as such, they were told only and vaguely that they were participating in a study of "how people understand and communicate math", and given no instructions beyond a request to stay within the camera's viewport. Furthermore, their task was simply to work on their homework in a self-selected group of 2–3; thus they are seen engaged in activities that they would have performed in any case (though perhaps not so consistently in groups). In all cases I had connections to the subjects outside of filming, in most cases because I was also enrolled in the class whose homework they worked on. One consequence is that they were generally quite comfortable around me and the camera, and their participation in multiple sessions over the course of a semester furthered this comfort. Another consequence is that I generally have access to a great deal of their work's context, having attended the same lectures, read the same textbook, and solved the same homework problems; in addition to this I collected copies of their scratch paper, assignments, lecture notes, and so forth whenever possible.

In presenting data, I follow the McNeill Lab convention of using [brackets] to indicate the full gesture including pre- and post-stroke holds, with **bolding** to mark the stroke itself (McNeill 1992). In addition, I use underlining to mark periods in which the speaker

is writing or drawing simultaneously with their speech. "<XXX>" indicates inaudible speech; "<X" and "X>" bracket places where the speech is almost inaudible and the transcription uncertain. Descriptions of the gesture or drawing are written underneath the appropriate line, aligned with the corresponding speech. Speakers are labeled 'L' and 'R', corresponding to their position on either the left or right in the clip at hand (from the point of view of a watcher looking at the videos); this uniquely identifies speakers in all examples given. Occasional examples from other sources label speakers 'A' and 'B'.

3 Speech, gesture, or...?

While the evidence is mounting that speech and gesture are at least closely connected, few theories would classify, e.g., drawing pictures as anything other than firmly extralinguistic. However, in the math videos there is a great deal of interaction between speech, gesture, and *ad hoc* features of the environment; whatever mechanisms support speech and gesture, they are very amenable to interaction with other aspects of cognition. In the recordings people are constantly writing, drawing, gesturing over drawings, pointing between written expressions, and so forth. It is generally impossible to understand the conversations from the sound alone; it would often be impossible to understand the conversations even after gesture was added if one could not also see the physical environment in which they work.

Additionally, the line between gesture and other things often becomes fuzzy; for instance, how does one make a principled distinction between gesturing with a pen one centimeter above a piece of paper, and actually marking the piece of paper with words or

diagrams?

In this section, I will present data showing just how blurred this linguistic/extralinguistic line can get, demonstrating that the mechanisms of speech and gesture are able to act in fine-grained online coordination with other mental processes.

3.1 Modality switching

"Backchanneling" is the communicative activity listeners engage in without actually taking the floor; common examples would be interjected *yeah*s, grunts, nods, and so on (Yngve 1970; Drummond and Hopper 1993). Such signals are used to show attention, request clarification, and generally maintain discourse cohesion; in the usual case they are done automatically and without conscious thought. In (1), the woman on the left starts a sentence vocally, *so if I have* and then continues it on paper, writing with her pen. Her interlocutor on the right maintains attention, and after some seconds of the writing have passed does a standard acknowledgment, *mmhmm*.

(1) **L** so if I have. . . mm. . . . . .
writes for 11 seconds

R mmhmm
8 seconds in, Figure 1

Clark (1996, 247) has observed that when responding to spoken language, such acknowledgment is generally given at the end of the "clause or phrase" being acknowledged. Here, obviously, there are no clauses, but the *mmhmm* occurs at the closest analogue: just as L is finishing one formula and moving her pen to start another. The conversation has transparently switched modalities while preserving the interactional properties

Figure 1: Example (1): mmhmm

of the discourse, and does so again; after she finishes writing, she looks back up at her partner and smoothly continues vocally.

Also of note about this snippet is the *mm*. As described by Clark and Fox Tree (2002), such words are used by speakers to signal the onset of an expected delay — and indeed, there is a pause of the appropriate length before she begins writing. A skeptic might hypothesize that upon encountering a concept difficult to express vocally, she said *mm* to gain time, struggled for some moments over how to formulate her next utterance, and then decided to resolve the difficulty by writing rather than speaking. However, the data contradicts this interpretation. During *so if I have* she directs her gaze at her interlocutor, then looks down at her paper in preparation to write; it only after she has completed this preparation and paused for a moment that the *mm* is uttered. Such delay management is almost entirely unconscious — only linguists know consciously the appropriate

usage of *uh* and *(u)mm* — and their use must necessarily be closely integrated with speech planning (Clark and Fox Tree 2002); that this usage is retained after the modality switch provides further evidence that she has not stopped one activity (speaking) and started another (writing), but rather changed modalities while maintaining a single interactional communicative activity.

3.2 Environmental anchoring

In (2) we see an example of environmentally anchored gestures. The student here is working on the homework for a class whose professor had a tendency to pause in the middle of lecturing and assign *ad hoc* problems related to whatever his current topic was. The speaker here is attempting to clarify his understanding of the problem statement while holding the notes he took earlier during lecture; the professor used a number of symbols without explanation in stating the problem, and the speaker is explaining that he assumed those to have the same meaning as when used earlier in the surrounding lecture. The top part of the page contains notes on the lecture content, and the very bottom is where he wrote the homework problem; he gestures over the appropriate portions of the notes to refer to each. This example will be analyzed further in §4.1; for now we note simply the way a pre-existing object integrates itself naturally into the discourse.

(2) **L** So I assumed that whatever we have [**up here**, omega star],
left hand sweeps up top of page, Figure 2
we [**can** use it here]
left hand forms pincer around bottom lines of page, Figure 3

Figure 2: Example (2): up here

Figure 3: Example (2): can use it here

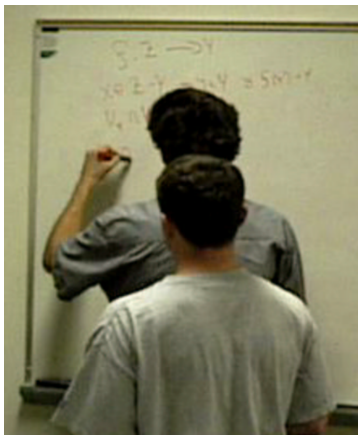

Figure 4: Example (3): V x intersect Y

3.3 Writing with and for speech

In (3), the speaker is explaining to his group some assumptions they can make about the sets that appear in their problem. In doing so he writes out expressions to represent the sets, and then points at and between the expressions to refer to them and their relationships.

(3) **R** so then we do $\underline{V}$— if $\underline{V\ x\ \text{intersect}\ Y}$
writes "$V_x \cap Y$" co-timed with words, Figure 4
if [**that's empty**… then we're done]
points at "$V_x \cap Y$" on board with right hand index finger, Figure 5
so, we're gonna assume it's not empty
similarly, so, $\underline{V\ y\ \text{intersect}\ Y}$, well that's—
writes "$V_y \cap Y$", again co-timed
if it's [**all of** Y], we're also done
points at "$V_y \cap Y$" with marking pen in left hand, Figure 6
[**by the same reason**]
right hand index finger points back and forth between "$V_x \cap Y$" and "$V_y \cap Y$"
cuz if [**this** is all of Y]
right hand index finger points at "$V_y \cap Y$"
then [**this** is empty]…
right hand index finger points at "$V_x \cap Y$"
cuz [**these are disjoint**]
On these, *each index finger points at one of the expressions, then both move up diagonally to previously written statement "$V_x \cap V_y = \emptyset$", reaching it on* disjoint, *Figure 7*

There are two things to note in particular about this sequence.

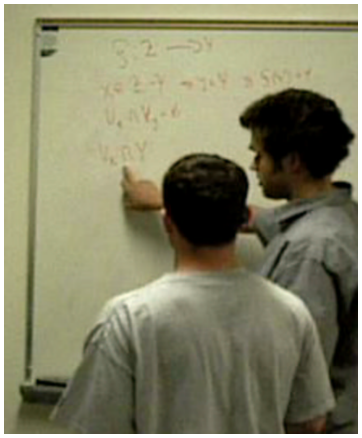

Figure 5: Example (3): that's empty

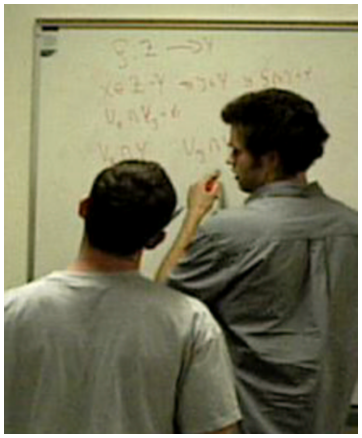

Figure 6: Example (3): all of Y

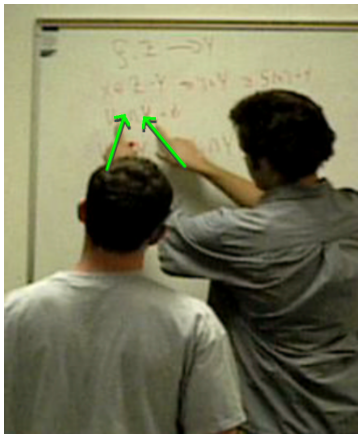

Figure 7: Example (3): these are disjoint

Firstly, the writing and the speech are co-timed precisely during the writing of $V_x \cap Y$ and $V_y \cap Y$; the speech slows and inserts pauses to ensure that each word is spoken while the corresponding symbol is being written, sacrificing fluency to preserve the co-timing that is normally considered a hallmark of co-speech gesture — this despite the fact that by most accounts, writing mathematical formulas should be entirely unlike gesture.

Secondly, observe how the writing and gesture interact; it is as if the speaker is writing things so as to have something to point at, as he first writes an expression, then points at it, then writes another, points between them, and then gestures their relation to a third proposition, represented by yet another written expression. The points themselves are not entirely transparent (after all, he is not actually intending to indicate pen marks *qua* pen marks), but they function analogously to those described by many others (Haviland 1993; Hindmarsh and Heath 2000; LeBaron and Streeck 2000; Goodwin 2003), and the general mechanisms that allow this will be analyzed further in §4.1. Here we are more interested in how the writing and gesture trade off; the evidence is only suggestive in the present case, but it seems not unreasonable to analyze the writing phases as 'preparatory strokes' for the gestures (and cf. example (10)). Furthermore, the speaker's next action after the transcribed portion is to reach over and name one of the sets by writing in an equals sign and a variable name. This is unlike conventional writing of, say, prose, in which one continuously extends a text in a basically linear manner; instead, he opportunistically builds new structure off of old structure, in a way far more reminiscent of his earlier *these are disjoint* gesture.

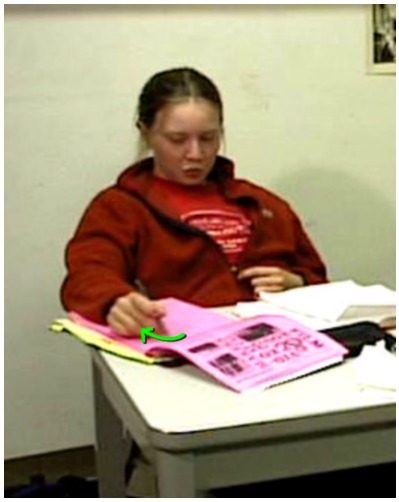

Figure 8: Example (4): pshew

3.4 Gestural drawing

In (4) we see something that blurs the line between gesture, writing, and drawing to invisibility — what one might call 'gestural drawing'. Here the speaker starts with markings on her paper corresponding to a group T that is 'acting' on a point. Without getting into the mathematical details, the idea of a group acting on a point is that each element of the group somehow modifies the point, so that from one point arises a whole collection of differently modified points.

(4) **L** you're taking it [**up…**]
draws an arrow showing the movement of a point

R you're moving it somewhere else

L acting on it with T, [**…pshew pshew pshew pshew pshew pshew**]
repeated strokes across paper from T to point, Figure 8
and then bringing it back.

This is represented by drawing multiple lines from the place corresponding to T towards the place corresponding to the point. These lines are drawn in a very fast and haphazard fashion, which makes no sense in terms of drawing, for a line is the same whether

drawn quick or slow — but if one conceptualizes the process as a series of group elements hurtling out of T and smacking into the point, then the motion does code appropriate force dynamics and imagistic structure. Furthermore, they are drawn on top of or nearly on top of each other, which again makes little sense if one's motivation is simply to draw a particular picture, but does accurately represent the motion of multiple elements whose paths share start and end points. Additionally, it is even unclear whether this is drawing or writing, since arrows are such a common conventionalized mathematical notation, and notational arrows can participate in all the normal things arrows can do in diagrams. To top it off the speaker accompanies each stroke with a vocal gesture in the form of a sound effect *pshew*, another case of co-timing between speech and drawing.

4 Persistent structure

Mental spaces theory (Fauconnier 1985) is a model of active conceptual structure — the things one is thinking about as well as the relations between them — that is especially useful in analyzing the ebb and flow of discourse reference. In normal conversation, referents and spaces are created extremely quickly, as many as several a sentence, and they generally disappear just as fast; humans only have so much working memory, and spoken language with its paucity of pronouns limits the number of entities available for reference at any particular time. The Access Principle (Fauconnier 1997, 41), which allows multiple entities to be referred to metonymically by a single term, helps somewhat, but it also introduces ambiguities. Most of the time, none of this causes any problem; people don't really need to pick out and refer to arbitrary bits from minutes earlier in a conversation,

and simply letting old mental spaces and referents drop out is unproblematic.

There are cases, however, where this is not true; else context and shared ground would hardly exist. It seems a useful enterprise to study in general the mechanisms by which speakers maintain conceptual entities and structure in their discourse, the purpose of such maintenance, and how the mechanisms and purposes interact. This section makes a start at such a study, examining persistence achieved through real space blends and repetition, and in particular argues that this is an area where principled trade-offs are made on-the-fly between speech, gesture, and other modalities.

4.1 Real space blends

Reality is an illusion, albeit a very persistent one. — Albert Einstein

The notion of real space blending, first introduced in the context of sign language research (Liddell 2003), is a powerful tool for the analysis of gesture. The idea is that one's physical surroundings have an ever-present conceptual representation, and therefore can participate in conceptual integration (Fauconnier 1997; Fauconnier and Turner 2002); such blends attach meaning to spatial locations and structures. Real space is different from other mental spaces though, in some key ways: it is shared, it continues to exist without attention or rehearsal, and it remains the primary domain for our faculties for spatial reasoning and action. Sharedness makes real space blends useful for communication; the other properties make them ideal for anchoring structure that is meant to persist for some time.

However, it is not only gesture and sign that involve the creation of real space blends;

we live our lives immersed in them, as we maintain constant awareness of environmental affordances, social spaces, the meaning of physically instantiated symbols, and so forth. There are times when such a blend is created and used entirely within a conversation, as in the canonical sign language examples or (10) below, but in other cases gestures may build off these pre-existing networks. For example, Hanks (1990) discusses the effect of context on pointing gestures, Goodwin (2000) and Özyürek (2000) the way gestures interact with the presence and location of interlocutors, Haviland (1993) and Levinson (2003) (among others) the effects a constant awareness of cardinal directions has on gestures generally, and so on; in this analysis, these are all instances of gestures interacting with pre-existing real space blends, and the same applies to Hutchins and Palen's (1997) analysis of "multilayered representations", Haviland's (2000) "laminated spaces", and so on.

Such an example in this paper was (2), with its gestures over class notes; the analysis is that the notes anchored a real space blend in which particular areas of the paper became integrated conceptually with the content of the writing in that area, allowing gestures over different areas to refer to different content. This cannot be analyzed as some sort of simply metonymic pointing, because there is more structure involved than a simple correspondence between spatial loci and conceptual entities. Consider the speaker's vertical sweep of flat fingers over the top of his paper: a similar horizontal sweep would have seemed very odd, because the conventional order of writing from top to bottom gives the blended space a preferred orientation for motion — despite the complete lack of spatial structure inherent in his ultimate referent, i.e. the professor's lecture. Similarly, the following vertical pincer would have been very odd as a horizontal pincer, because the writing has a conventional width that does not need to be specified, and thus verti-

cally positioned delimiters are both necessary and sufficient to pick out a subregion of the blended space. This is evidence that this is truly a blend, with concomitant emergent structure.

Real space blends are a powerful way to create persistent structure, because they can draw on the special persistent properties of reality. However, not all such blends are created equal; there are many methods of creating them, and systematic differences that can make one or another appropriate to a given situation. Some key axes of difference include the degree of sharedness, the total amount of structure encoded, the ease with which the resulting blends persist, and the amount of preparation required.

On one end of the scale we have *private blends*, by which I mean a real space blend formed inside one's own head without any external indication; for example, when trying to remember a person's looks, one might mentally 'project' an image of their face onto a blank wall. This is entirely unshared, and may be easy to forget; the amount of structure encoded can vary, but matters little for communication, since it is not shared. On the other hand, such a blend requires no preparation at all; one can do it at any time and on a moment's notice.

One might follow up such a blend by beginning to gesture in it — consistently using different hands to gesture about contrasting concepts, for instance. This is still a bit shaky in sharedness; while one's interlocutor may pick up on the blend from just this, one cannot depend on it. In most such cases the structure encoded is rather small — two contrasting locations, in the example; without further cues, gestures that depended on the details of a blend containing complex spatial structure would lapse into incomprehensibility. The preparation required is again minimal, but with no spatial reminders of

the blend's existence, it can easily lapse from memory — it does not persist very well.[1] Therefore, we refer to these real space blends as *unanchored*.

A solution to this lack of persistence is to create reminders; a gesturer can leave their hands held in a single place, thus changing the structure of real space to better support their blend. We call such a blend *gesturally anchored*. One can rely on these being shared with one's interlocutor, they can be at least somewhat *structured*, i.e., contain some structure in the hand shape and position that is mapped to the internal structure of the source space — though the amount and type of spatial structure available is limited by anatomy. In other cases the mapping is cruder, creating gesturally anchored blends that are relatively *unstructured* — simply two hands held up to indicate two distinct objects, for instance. Again, such blends require no preparation ahead of time. They do cause a problem, though; hands are a limited resource. A hand that is fixed in place maintaining a blend is a hand that is severely limited in its capacity to perform further gestures or anchor other blends.

If one instead blends with external artifacts to create what I will denote an *artifactually anchored*[2] blend, this last problem is solved, but at the cost of either preparation or structure. If one has not prepared in advance, one is restricted to whatever objects happen to be around, and they are unlikely to have much inherent structure useful for a blend; pennies on a table can be used to represent relative locations of people at a described party, but provide little support for then describing clothing styles, facial expressions, or even indicating the direction each person was facing. On the other hand, one can cre-

[1] At least for speakers of vocal languages. Signed languages have grammaticalized this use, and their users tend to have a great deal of practice at retaining the meaning of loci. Gaining this ability is one of the more difficult things about first learning a signed language.

[2] Or *materially anchored*; see §7.

ate artifacts specialized for supporting highly structured blends, like the models used in presenting architectural designs. Such artifacts are extremely powerful for talking about a particular topic: the blend they anchor is generally available to anyone who looks at them even without help from the speaker, and they require no maintenance once created, making them extremely persistent — a speaker might relax their hands and lose a gesturally anchored blend, but a physical object will remain until actively removed. Unfortunately, they take so much work to create that they are useless for most everyday conversations, and are quite inflexible to boot. A speaker can modify a gestural blend on the fly as the conversation shifts; a pre-existing artifact usually cannot be so adapted. Hoque (2003, see also LeBaron and Streeck 2000) observes this as well in her analyses of architectural design reviews; speakers will often gesture around their models, but will also pull their hands up off the model and gesture above it when explaining aspects that the model's solidity make difficult to otherwise express. Interestingly, in such cases they do not abandon entirely the blend their model anchors; their off-model gestures tend to preserve the model's shape, orientation, and general size.

However, there is a partial solution to the difficulties presented by inflexible, high-preparation anchors. Drawing and writing are ways of creating custom artifacts that can support very complex structures, are fully persistent and shared, leave the speaker's hands free, and yet require relatively little preparation; if one needs to build a lot of persistent structure but is unable to prepare in advance, pencil and paper make ideal tools.

These are the kind of principled differences one can make between different techniques; note that the dimensions along which we compare each type can be applied equally well without regard to modality, and they determine the relative utility of differ-

ent methods of creating persistent structure in many situations. It is not uncommon, for instance, to see someone first attempt a complicated explanation with gesture, and then switch to diagramming when the need for structure and persistence become too high. Likewise, it is no coincidence that most lecturers prefer to use slides; giving a prepared lecture is a situation where the downsides of structured artifactual blends do not apply. This approach provides a unified way to analyze gesture, writing, object manipulation, drawing, and so on. This unification is promising firstly because as we saw above, the boundaries between the phenomena are often very unclear, and so our theory should not predict each to be entirely different; on the other hand, it allows us to describe why we saw the differences we have. Later examples will demonstrate subjects whose behavior is best explained by reference to the distinctions given above.

4.2 Repetition

Another way of maintaining persistent structure in discourse is to use repetition and rephrasing of words, phrases, or syntactic constructions. This occurs in both speech (Halliday and Hasan 1976; Tannen 1989) and gesture (Tabensky 2001). Some examples of this repetition, both between interlocutors and within a single speaker, are:

(5) **A** If I were you, I'd run.

B If you were me, you'd be good looking.

(Six-String Samurai)[3]

(6) If you buy SCO's argument that Linux *is* Unix with the serial numbers filed off, then SCO might actually have a leg to stand on here. If, instead, you believe that Linux is Linux and SCO has no right to steal it, SCO's non-compete argument makes no sense. (Linux Weekly News, `http://lwn.net/Articles/58921/`, emphasis original)

[3]See also (Goodwin and Goodwin 1987) for similar cases occurring in children's playground arguments.

The point here is that not just that the constructions *If I/you were you/me, I/you'd X* and *Linux is X* are repeated; that alone would merely be evidence of syntactic priming or similar. The important thing is that this repetition has semantics; by using repeated constructions the speakers suggest a comparison between the meaning of the different instances.

In (5), the second speaker pokes fun at the first, and deflates his threat, by using his own form to mock him. Notice that the repetition is not purely syntactic, because of the switch in deixis. From a mental spaces viewpoint, such conditionals work by first creating a premise space (in this case a blended space in which A is somehow identified with B), and then describing consequences in this space (Fauconnier 1997). The switch in deixis means that B is not just echoing A's construction, but explicitly retaining the same referents and reinvoking the same premise space, and then offering an alternative consequent. That the same space is retained is important in achieving B's communicative purpose.

Firstly, it makes his response a contradiction. B is engaged in responding to a threat with exaggerated unconcern. Had A expressed the threat differently, instead saying e.g. *You had better run*, then B's response would have had the opposite effect, becoming a boorish insult and implying that A had successfully rattled him.

Secondly, B is deflating his interlocutor with humor, and the humor also arises from the retention of a single premise space. While both statements involve a blend in which A somehow becomes B, the structure of these blends is different. In the first statement, we are to infer that if A, with A's knowledge, were in B's situation, then A would run. The second statement asks us to imagine that A were somehow possessed of B's physical body. (Lakoff 1996; Fauconnier and Turner 2002) While both are described as A 'being'

B, the details in each case are quite different, and this is the source of the humor. Simply describing two different situations is not funny; but the sudden reconstrual of a single situation in a radically different way is (Coulson 2001), and A uses this humor to make the threat seem ridiculous.

In (6), *Linux is Linux* does not mean any of the various things that it might in isolation; it is clearly intended to contrast in form with the line *Linux is Unix*, and thus one interprets it to contrast in meaning as well. *Linux is Linux* here means just what *Linux is not Unix* would, while additionally letting the speaker (who in fact considers SCO's argument absurd) frame his preferred interpretation as a logical tautology. In both examples, the second use would be a non-sequitur if the first had not occurred.

Another, somewhat different sort of repetition is demonstrated by the imagined exchange:

(7) **A** Maybe you should try being a GSI for a semester?

B Well, technically, I can't, not being a GS, but yeah.

Here the second speaker is taking advantage of shared knowledge that GSI is an acronym for Graduate Student Instructor; saying *GS* is a creative usage that would make no sense whatsoever in a neutral context, but the formal similarity to the earlier *GSI* is obvious, and this allows one to extract from *GS* its meaning "Graduate Student".

Similar things occur in gesture. Tabensky (2001) gives several examples of people picking up on interlocutors's gestures and repeating them either identically or with deliberate, contrastive modifications. The gestures others have analyzed as "catchments" (McNeill 2000a; McNeill et al. 2001) are in this reading essentially gestures of self-repetition, and unsurprisingly can show similar behavior. This sort of repetition can also be used with

vocalic gestures, as in the attested example (8), where speaker A suggests something, and speaker B gives his opinion that this suggestion is deeply unacceptable:

(8) **A** I'm just saying that I think it would be funny,
if Aragorn and Arwen were cast as Lord Azrael and Ms. Coulter.

B Oh, I see…
You don't mean fúnny funny.
you mean <guhh agh ghh> funny.
Choking noises and an extremely pained expression

Here speaker B comments on A's choice of the word *funny* by echoing it with modifiers. He does this twice, with repeating grammatical structure and stress patterns, contrasting two different 'meanings' of *funny* — one, the canonical meaning, indicated by just saying *funny* with a strong stress, and the second illustrated by a dramatic enactment of his reaction to the suggestion. Choking noises do not constitute a lexical item in English, and cannot normally participate in grammatical structures, but here the alignment with the previous sentence is used to coerce the choking noises into filling a constructional slot as if they were an ordinary noun.

I would also be unsurprised to see cross-modal repetition — one person gesturing a spatial blend and the other repeating it back in writing, or vice versa — but have not yet looked for such in my data. This would be an intriguing addendum to the sort of cross-modal interaction demonstrated in (1), and a good candidate for further investigation.

5 Persistent structure in discourse

Having examined several methods used to create persistent structure, it is time to consider the uses to which they are put, and make good on my claim that the choice of

mechanism is principled. There are a number of these uses in ordinary communication; I have to date identified three: topic holding, providing a substrate for further contrastive signals, and structuring discourse.

5.1 Topic holding

Hold gestures can be used to maintain a topic through attempted interruption. This works because gesturally anchored blends are persistent and shared, but not permanent; when one has blended topical information with space in front of one, and retains that blend obviously and effortfully, then it is clear to all parties that one intends to stay on that topic. Thus out of all the real space blending mechanisms described above, it is holds that one often sees used for topic maintenance; the other mechanisms are either non-obvious or non-effortful.

Another method to hold a topic through disfluency and attempted interruption is to repeat the last word one said; this signals clearly that one is not ready to give up the floor, and furthermore, that one was unfinished with the previous sentence and has more to say on the subject. Other pause fillers can signal the former, but are less effective at the latter.

Both of these techniques are used in (9):

(9) **R** [okay, umm, so he says basically you do that twice, for…for…
starts with hands down and splayed tensely, where they have been illustrating two sets

L I have another way of doing it

R for…for…
*looks over at **L**, hands still splayed*

L I have a different way of doing it

R aren't you special…
hands still splayed, Figure 9

Figure 9: Example (9): aren't you special

cuz this way of doing it is,
like, okay,
that's how you prove that this compact set,
and, if there's an open set around this compact set,
and there's an open set around this point,
and then he said you expand that point to a compact set—
whole sequence uses gestures with and around the two held hands, Figure 10

L I have a different way of doing it

R . . .][**. . .**]what's your way of doing it
drops both hands to lap, Figure 11

Here the woman on the right has for the last several minutes been explaining a long proof involving a point and a set and how the techniques used in that proof are relevant to their current homework problem. All through the explanation she has held her right hand to represent the point and her left to represent the set. Having finished the explanation, she gets stuck trying to describe its relevance, and starts repeating her last word, *for. . . for. . . for. . . for. . . .* The woman on the left takes this opportunity to request the floor

Figure 10: Example (9): cuz this way of doing it

Figure 11: Example (9): what's your way of doing it

with her *I have another way of doing it…I have a different way of doing it*. R is having none of it; she wants to finish her explanation, and so keeps holding her hands where they are and repeating herself, and then when the other woman repeats the request a second time, she rebuffs it with *aren't you special* (said teasingly; both women are close friends), and proceeds to finish her explanation, still using the held hands to anchor her explanation's gestures. This concluded, L repeats her request for a third time, and this time R allows it, explicitly giving up the floor with *what's your way of doing it* and finally dropping her hands. Though the still frame does not show it, this last is done after a moment of looking at the other woman, visibly considering, and then the dropping of her hands is accompanied by a visible relaxation of her whole body as she decides to cede the floor.

It is informative to contrast some aspects of these two topic holding methods. Verbal repetition is useful in that it establishes the speaker's continued desire for the verbal floor, which a static gesture cannot do. On the other hand, the limitations of auditory attention mean that only one person at a time can make effective use of the verbal stream, and this makes possible a common tactic: when two speakers desire the floor, each speaks over the other, until one is forced off the air (Clark 1996). Hold gestures are less effective for gaining turns, but they can support a simultaneous verbal strategy, and this gives an advantage over a pure verbal approach: hands are not a shared resource, and gestures cannot be drowned out by an interlocutor.

Also of note in this sequence is the repetition begun with *I have another way of doing it*. L repeats this phrase three times with small variation, making it clear that it is a single request being renewed, rather than some new request based on some new idea she has had. Then when R finally accepts the new topic, she uses a rephrasing *what's your way of*

doing it, making it clear that she is asking for the *way of doing it* mentioned in the earlier requests.

5.2 Persistent substrata

A common use of persistent structure in the math videos is to provide context for further signals, to create and hold a complex real space blend that speakers can then continue to gesture or draw over. (3) gave an example of this for writing, and (4) an example for drawing[4]; (5)–(8) show a similar effect using rephrasing. In this section, we analyze a purely gestural case that uses a real space blend: (Note subscripts are used to match up the opening and closing braces of two temporally overlapping holds)

(10) **R** so [$_1$**my** original closed set,
puts right hand up in air, holds, Figure 12
right, was [the comp**lement of some** open set.]
scoops left hand around to indicate open set, Figure 13
now I'm [taking **that** open set,]
repeats scoop, quicker this time, Figure 14
union [another open set. . . right?]
left hand hold, Figure 15
[**that open set**]
scoop again, Figure 16
union [$_2$**U**,
left hand hold again, Figure 17
and taking the]$_1$[**complement of that**
right hand moves rightward (woman's right), Figure 18
]]$_2$. . .
both hands drop

Here the woman is describing how she constructed a particular set and how she knows that it is "closed" (a mathematical property whose details are unimportant for our purposes), basing her construction and her reasoning on another set — her *original closed set*,

[4]If one can consider that to be an example of drawing, which is somewhat unclear. In any case, the phenomenon does occur with drawing.

Figure 12: Example (10): my original closed set

Figure 13: Example (10): complement of some open set

Figure 14: Example (10): that open set

Figure 15: Example (10): another open set

Figure 16: Example (10): that open set

Figure 17: Example (10): union U

Figure 18: Example (10): the complement of that

which is represented by her right hand held up in a cupped shape commonly used to represent sets. This hand stays up and motionless for the first part of the sequence as she explains the other sets involved by reference to the first. The first open set is described as the complement of the original closed set, and represented gesturally by a sweep around the right hand hold. The complement of a set is its 'reverse image' — all the points not in the set are in the complement, and all the points *in* the set are *not* in the complement — so this gesture indicating all the space around the held hand is a sensible representation. The second open set is described as some other open set named U, conceptualized as intersecting the original closed set; it is represented gesturally with her left hand partially overlapping her right hand. Then she describes the construction of the new set; the means of construction is essentially to take a U-shaped bite out of the original set, and this presents a quandary: she wants to move her right hand, to represent the new closed set,

but it is the anchor point for the whole scene. She is up to the challenge; she keeps her last left hand U gesture stationary to provide a new anchor, and moves her right hand to her right, so that her hands no longer overlap, just as her new closed set no longer intersects U.

The overall result is a clear sequence of gestures showing her construction and reasoning; none of which, however, would have made any sense without the held hands maintaining the blend and providing reference points. It also demonstrates how the properties of a speaker's blend anchor affects their actions; the tricky switchover in anchors is necessitated by the resource limitations created by use of a gesturally anchored blend.

5.3 Structuring discourse

(11) I have three things to discuss. Firstly, ... Secondly, ...

Usages like (11) are very common. The idea is to blend the temporal sequence of a discussion with the sequence of the counting numbers, mapping discrete topics to discrete numbers. A common accompanying gesture is to hold up three fingers, making use of a conventionalized mapping between fingers and counting numbers[5] and touching each finger in turn as each topic is discussed. Alternatively, one might hear the above phrase uttered by a lecturer standing in front of a slide with three topics outlined on it, who then pointed at each line as the topic was described verbally; here one uses the conventionalized reading direction of English to impose a linear sequence on lines of text and map them to the counting numbers. Or possibly the slide would be shown once, and then repeated throughout the lecture at the beginning of detailed discussion of each topic, with

[5]The details of this mapping are in fact culturally specific; different communities have different conventions for which finger to start counting from, which groups of fingers to use for particular numbers, etc.

the relevant portion highlighted.

Each provides a somewhat different example of persistent structure, created through speech alone, speech plus gesture, speech plus writing plus gesture, or speech plus writing alone, and maintained through short stretches by repeated verbal referrals or consistent physical presence, or maintained through the entire lecture by repeated showings. All serve the same communicative purpose of providing a skeleton to structure the discourse, despite using different modalities and mechanisms, demonstrating the unifying power of the persistent structure approach.

6 Case studies

Earlier examples were chosen to unambiguously demonstrate a particular point of interest with a minimum of distractions; in reality, however, most interaction is not so clean. I have therefore selected two longer, more complicated examples to demonstrate how much the various phenomena interweave, and how quickly speakers can switch between strategies. Of course, like any consciously selected sub-sample, these clips are still not really representative; for example, in the real math videos the subjects often spend minutes at a time simply staring blankly at their papers. Sequences of this sort are, however, still quite common, and are valuable to provide another point on the range of possible behaviors.

6.1 Spiral sequence

This example consists of three separate clips, all of which center around a particular drawing of a spiral, and a discussion of how to squeeze the whole complex plane down onto a line (and in particular the line the spiral represents). In the first clip, (12), the spiral is drawn. The accompanying speech identifies it as *the logarithmic spiral*, a particular curve in the complex plane; drawing it establishes that portion of the board as a surrogate for the complex plane, with the origin at the center of the spiral, and the actual line of course standing in for the abstract curve. He does not draw anything else, such as the coordinate axes that are conventionally used to represent the complex plane; the rest of the blend is implied by the simple line, and as we will see from their interactions with it below, it is sufficient to anchor the whole thing.

(12) **R** the logarithmic spiral...
draws a spiral on the board, Figure 19

Having drawn the spiral, he then ignores it; it is two minutes before they mention or interact with the spiral again, in (13). When they do, it is the same man who drew it in (12), talking and gesturing about circles, and how the points on each circle will map onto the spiral. He refers to *all the circles* three times. The first time, he gestures circles in front of himself in a standard iconic way. The second time, he says *e to the t e to the i theta are all* rather than *all the circles*, but this should still be considered a self-repetition, as *e to the t e to the i theta* is the equation for a circle, and it still contains the idiosyncratic use of *all*. This time he says it while writing the corresponding expression on the board, with co-timing as in (3) and drawing an arrow towards an earlier expression, building off this structure created earlier. The third time he says it, he has moved over to his spiral drawing, and

Figure 19: Example (12): the logarithmic spiral

uses his pen to trace a circle in the air over it. This last shows how the complete complex plane is accessible to gestures within the real space blend, though only the spiral is drawn. He then describes the point that each circle will map to, first pointing at the expression for them on the board, and then scribbling the same expression on the air with his pen, again co-timed at the level of individual symbols; this last is yet another place where the line between writing and gesture becomes decidedly murky.

(NB: I consistently refer to the man at the board as R and the woman as L throughout, though they reverse their positions for Figures 20–21. The other man does not speak or interact at any point.)

(13) **L** I'm thinking we're gonna send all the **[circles]**
both hands up and rounded to form circle, Figure 20
. . . like e to the t e to the i theta are all going to go to these
writes "$e^t e^{i\theta} \longrightarrow$", the arrow pointing at "$e^\theta e^{i\theta}$", the previously written equation for the logarithmic spiral, Figure 21

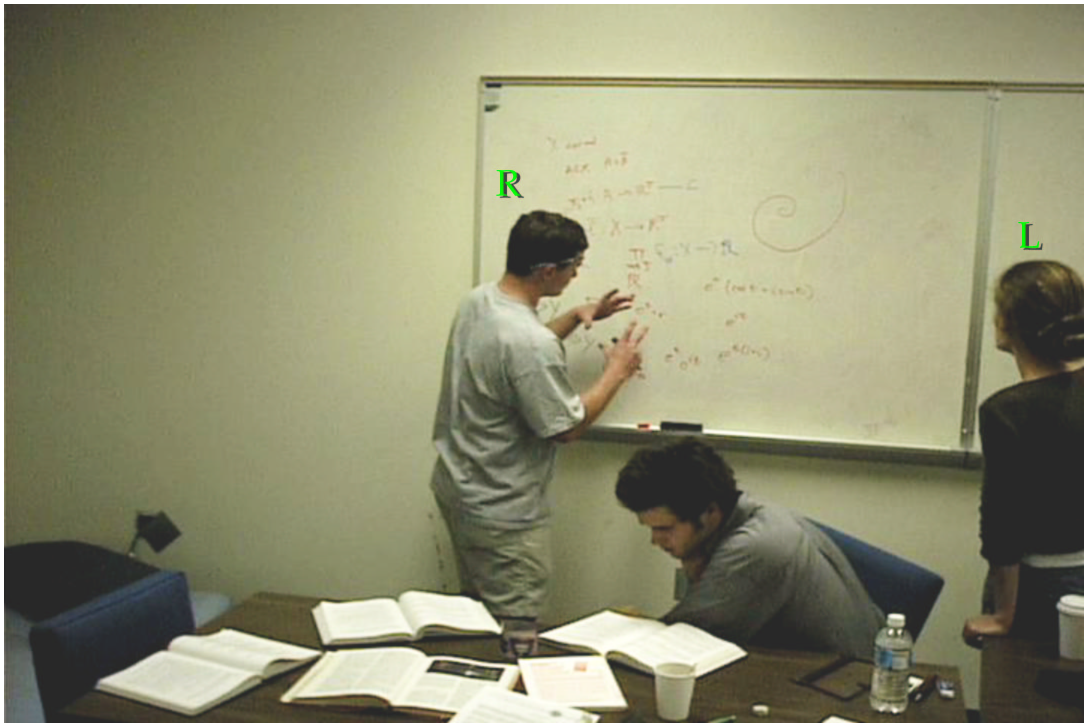

Figure 20: Example (13): all the circles

you know what I'm sayin?
moves to other side of board
like…all the circles[…]
traces a circle on board over spiral with pen, but doesn't draw, Figure 22
are gonna go to[…]the point on the circle that
points at "$e^{\theta}e^{i\theta}$", Figure 23
is [**e to the theta times e to the i theta**]
sloppily writes "$e^{\theta}e^{i\theta}$" in air, Figure 24

Thirty seconds later, after some diversion, they return to discussion of the spiral, again with the same person speaking. His first gesture is a point at the center of the spiral while talking about the complex plane's origin; recall that when the whole blend implied by the spiral's presence is taken into account, its center corresponds to the origin. He then repeats the phrase *all the circles* for the fourth time, and this time draws an actual circle on the board, using the same motion as his previous gesture, but this time with the pen in contact with the board.

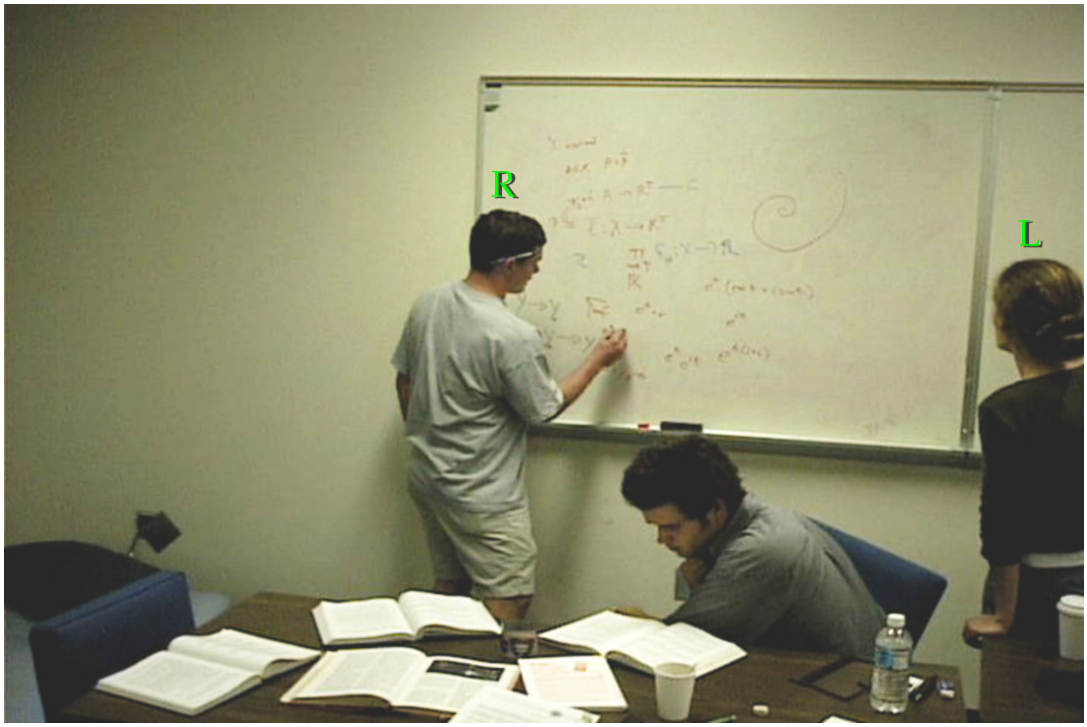

Figure 21: Example (13): e to the t e to the i theta

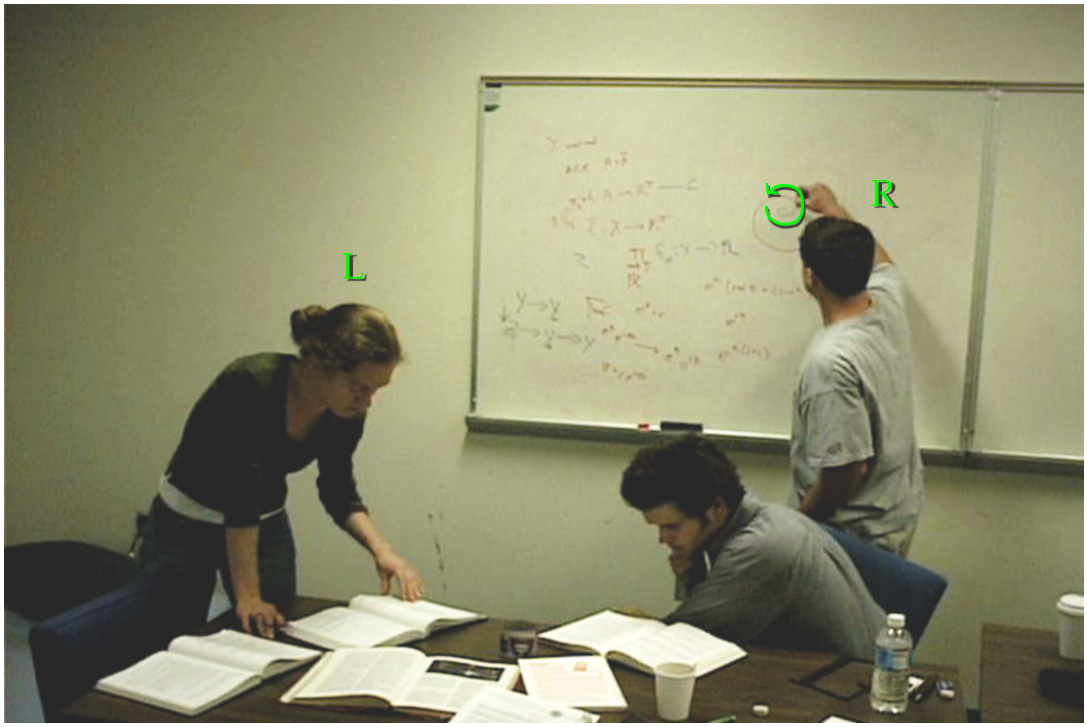

Figure 22: Example (13): all the circles

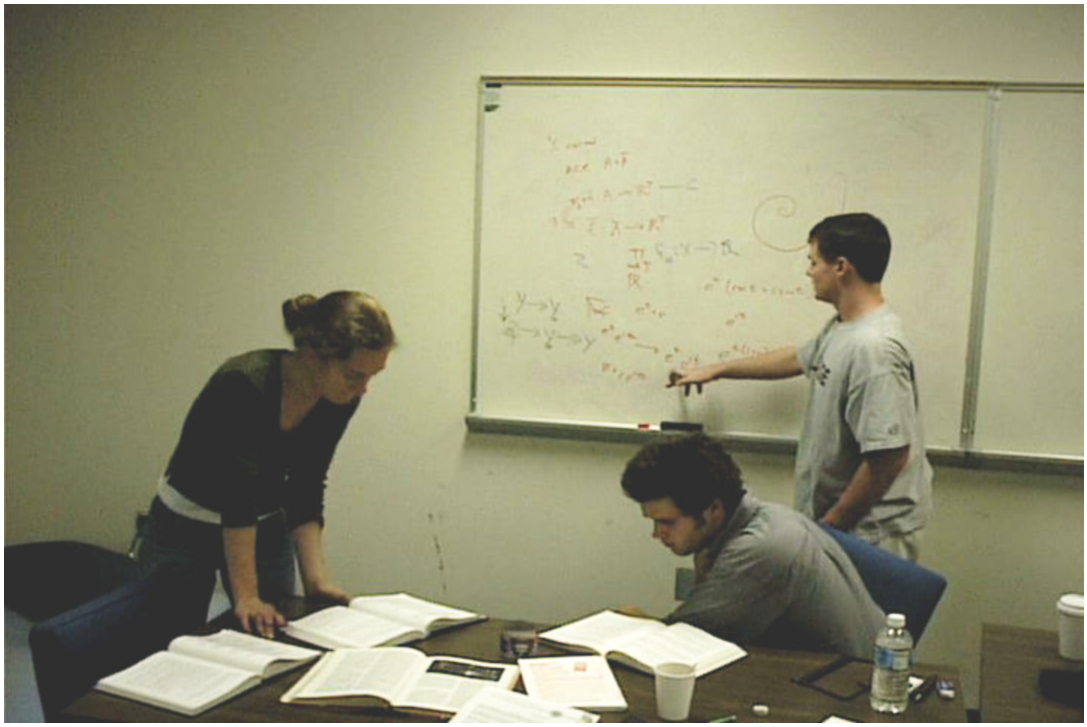

Figure 23: Example (13): to the point

Figure 24: Example (13): e to the theta

At this point, the woman interrupts to give an objection to his proposed function. She walks over to his drawing, and begins gesturing over it; first running her hand down one arm of the spiral while making a statement about the points on the spiral, and then making pinching motions on both sides of the spiral to describe the behavior of points off the spiral as the function pulls them towards it, showing that the entire complex plane is accessible to her gestures as well. She can use his diagram as a substrate for making her own statements, in a way reminiscent of what Furuyama (2000) refers to as "collaborative gestures", in which one person reaches into another's gesture space to manipulate the structure of their blend directly.

The man then clarifies his original description, with another repetition *all of the circles* accompanied by a gesture over the circle drawn on the previous repetition, and again he follows this by a reference to the point of intersection between the circle and the spiral. This time, however, there is actually an intersection drawn on the board for him to point at, and he does. Until the circle was drawn, this was not possible; adding to the diagram gave him additional affordances for gesturing.

Note that all these repetitions of *all the circles* serve a purpose; each is co-timed with some action that represents the circles in a different way: firstly with a standard iconic gesture, secondly with a mathematical formula, thirdly with a gesture over his diagram, and fourthly with the drawing of a circle. By repeating the same speech with each, he ties together these disparate representations. This is similar to a phenomenon observed in architectural reviews; the presenter there has many artifacts — plans, pictures, models — all representing the same building, and while presenting will often repeat a gesture over multiple such representations (Hoque 2003). This is almost identical to what R is doing,

but with different modalities; it uses repeated gestures to tie together artifactual representations rather than repeated speech to tie together gestural representations. On the other hand, there are important differences; gesture is better able to represent spatial information than speech, and the presenters take advantage of this. Their detailed gestures over multiple objects demonstrate how to align the internal structure of each representation with the internal structure of the others, while co-timed speech is only able to indicate general coreference.

Finally, R turns away from the board to finish his explanation, and does an iconic spiral in front of himself co-timed with *spiral*, in a manner very similar to his first 'circle' gesture.

(14) **L** the pre-image of the [**origin can't just be** the] origin
points with pen at center of spiral on board, multiple beats, Figure 25
. . .
so that can't be right, like
we can't– [we can't be sending **all the circles**] to–
draws a circle on the board, along same path as had gestured a circle earlier, Figure 26

R cause it seems like you're like [nailing down **all this stuff here**
O hand taps down along spiral, Figure 27
and then you're gonna have to **break something apart too**] <XXX>
hand stops moving, uses fingers as pincers to "pull"
bits of space toward the spiral, Figure 28

L well I was thinking like we could send [all of the **circles**]
index finger traces previously drawn circle, Figure 29
[to the unique **point** that this thing intersects]
index finger stops on intersection of drawn circle and drawn spiral, Figure 30
cuz [the **spiral**] intersects a circle at one and only one point
traces spiral in air in front of him with index finger, Figure 31

This whole sequence centers around a single persistent real space blend created by drawing; this blend persists for minutes on end despite distraction, warps the space around it to contain a whole complex plane, serves as a shared locus of further gestures

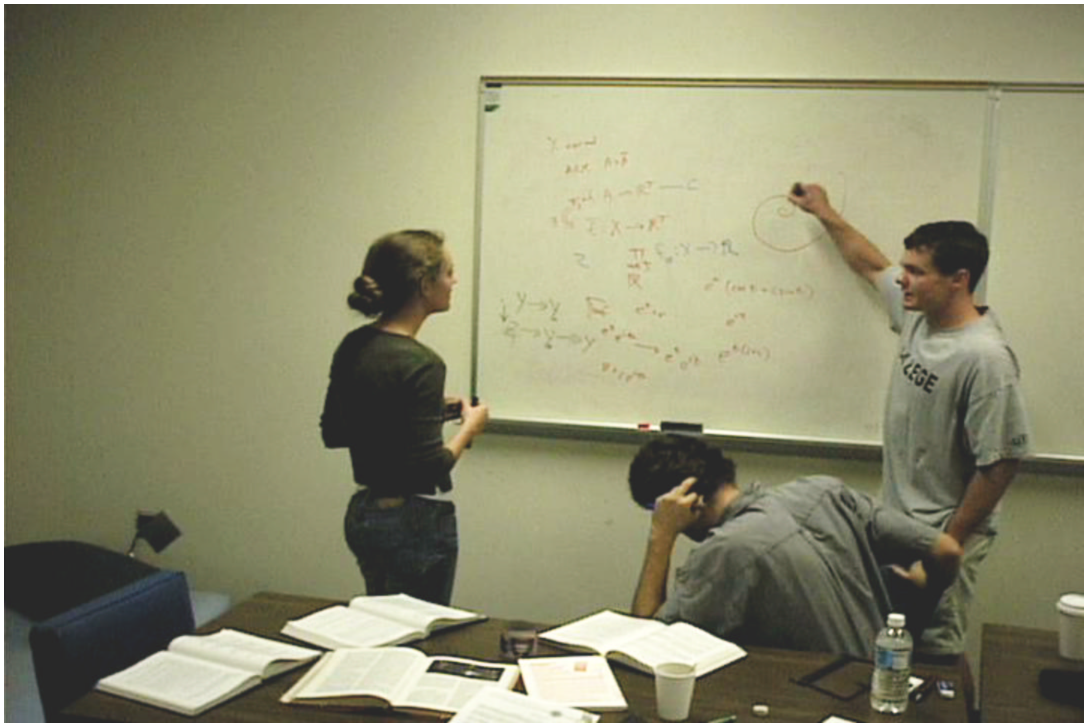

Figure 25: Example (14): the origin

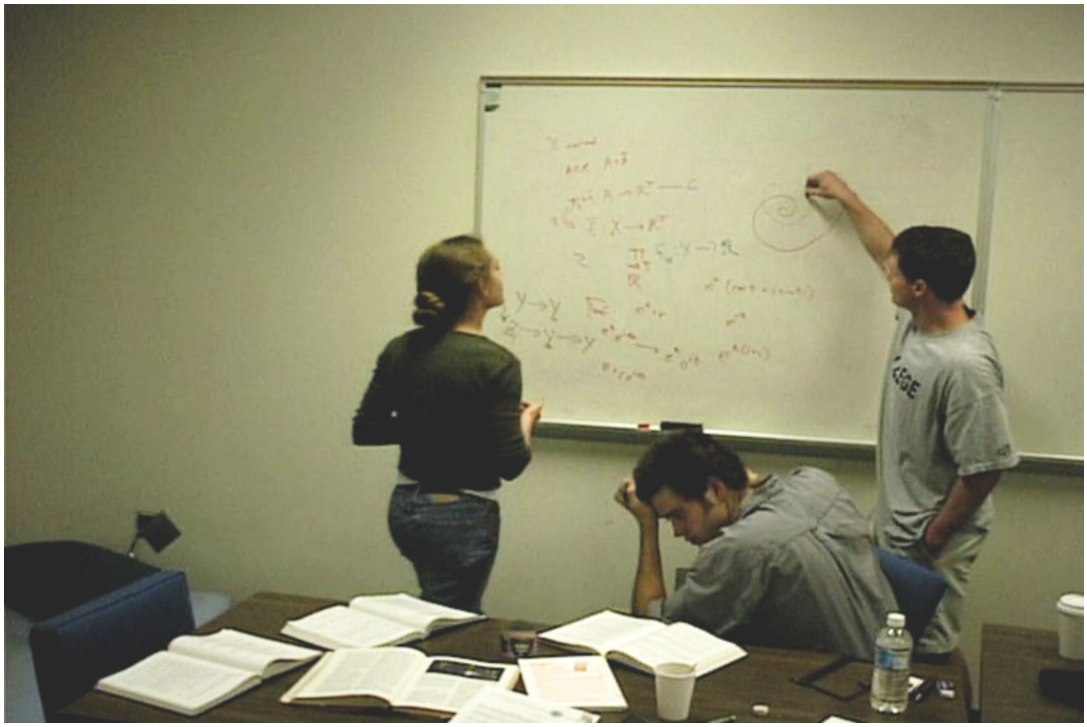

Figure 26: Example (14): all the circles

Figure 27: Example (14): all this stuff here

Figure 28: Example (14): break something apart too

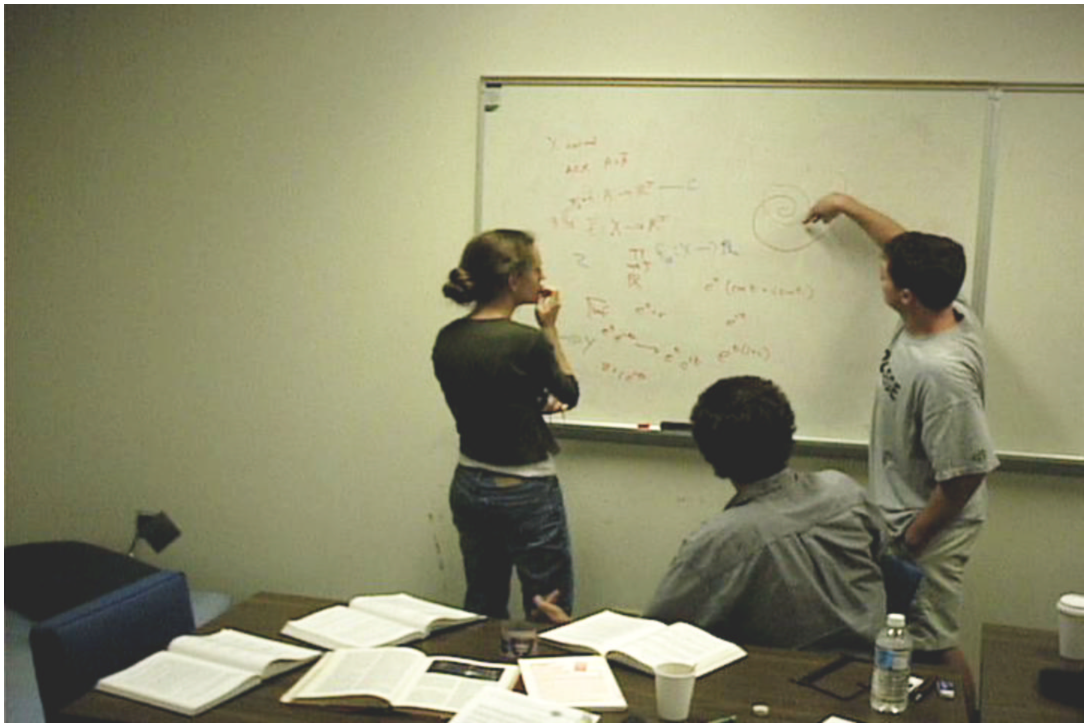

Figure 29: Example (14): all of the circles

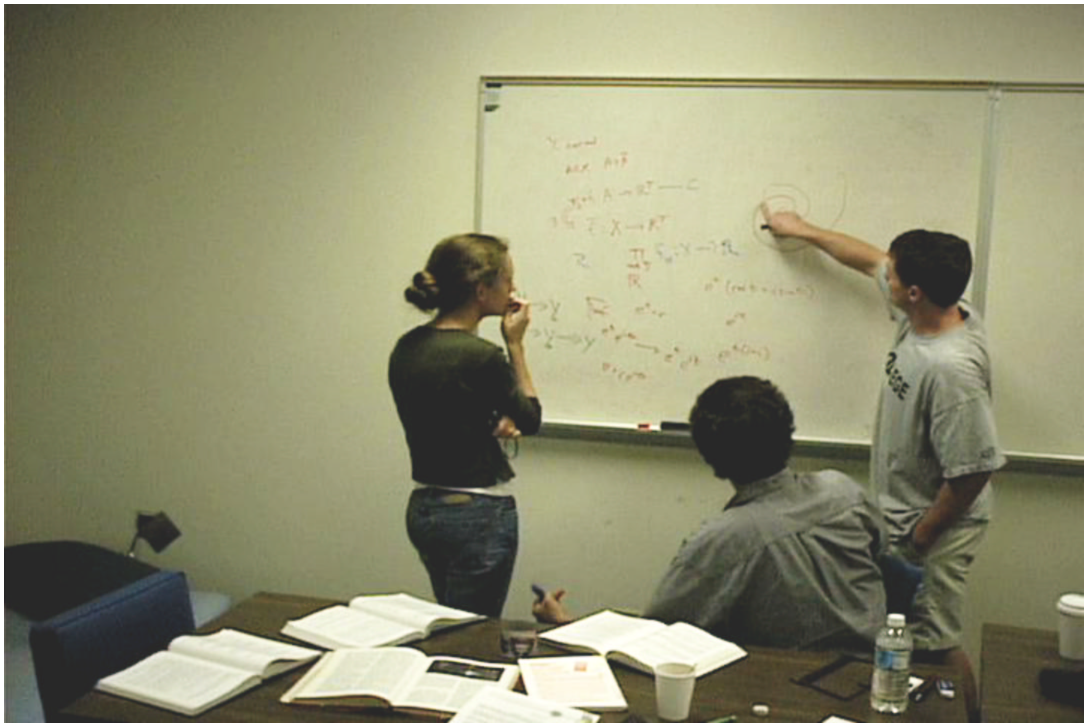

Figure 30: Example (14): the unique point

Figure 31: Example (14): the spiral

for two different people, and is further modified on the fly to create additional affordances for additional gestures. Other gestures, writing and drawing are used during the same period to refer to the equivalent objects, and these multiple representations are tied together by verbal repetition.

6.2 Primitive nth roots of unity

In this example, R tries to explain his idea for a proof involving "primitive nth roots of unity". (The only fact about primitive nth roots of unity that is relevant here is that they form a subset of all the nth roots of unity.)

(15) **R** it can only contain finitely primitive nth roots of unity, for different n, right
[take all those, right
forms a two handed cupped hold, Figure 32
so you have— you have a set **of all the primitive roots** of unity it can contain
two hands beat twice

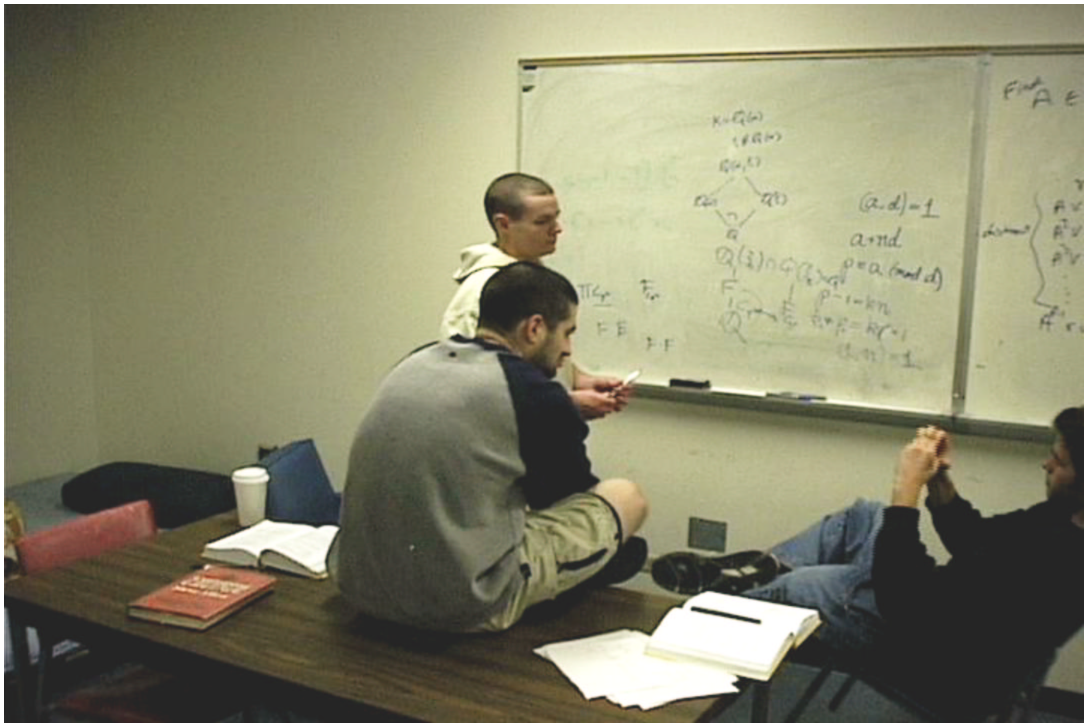

Figure 32: Example (15): take all those, right

L or all the roots of unity [**it contains**] just, period
two hands make loose enclosure, Figure 33

R well, b— but the primitive ones are the ones **that like**
more beats of two hand hold, Figure 34
matter for this case
][. . . **then** for any zeta]
points at diagram on board containing "ζ", Figure 35
umm. . .
[**relatively** prime to that]
repeats point at board
[**to the— to the set of zetas**
reforms two-hand hold, more beats, Figure 36
. . . so it's relatively prime to each one of those]
left hand stays fixed, right hand makes repeated jabbing points around it, Figure 37
then by like induction on [**this**] theorem
jabs open page of textbook with left hand, Figure 38
[**that** intersection] will be $\mathbb{Q}$
points at board again, Figure 39

To start with, he establishes that the set of such roots is finite, and then invites listeners to form the set of them all; this set is represented by a two-handed container hold in front

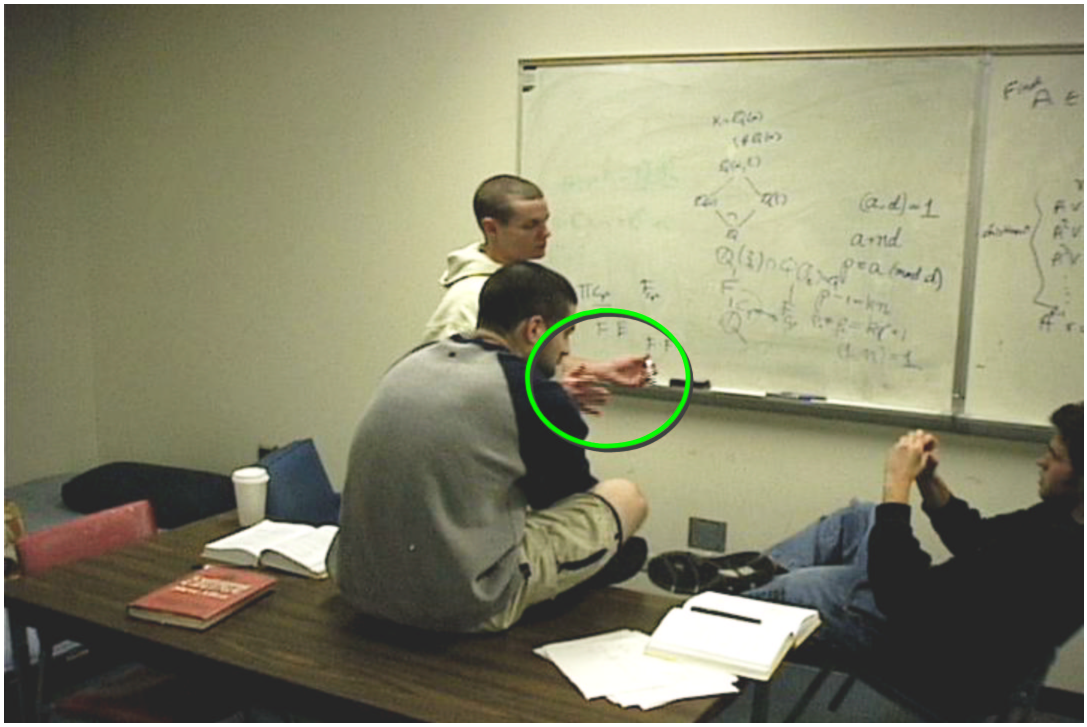

Figure 33: Example (15): or all the roots of unity that it contains

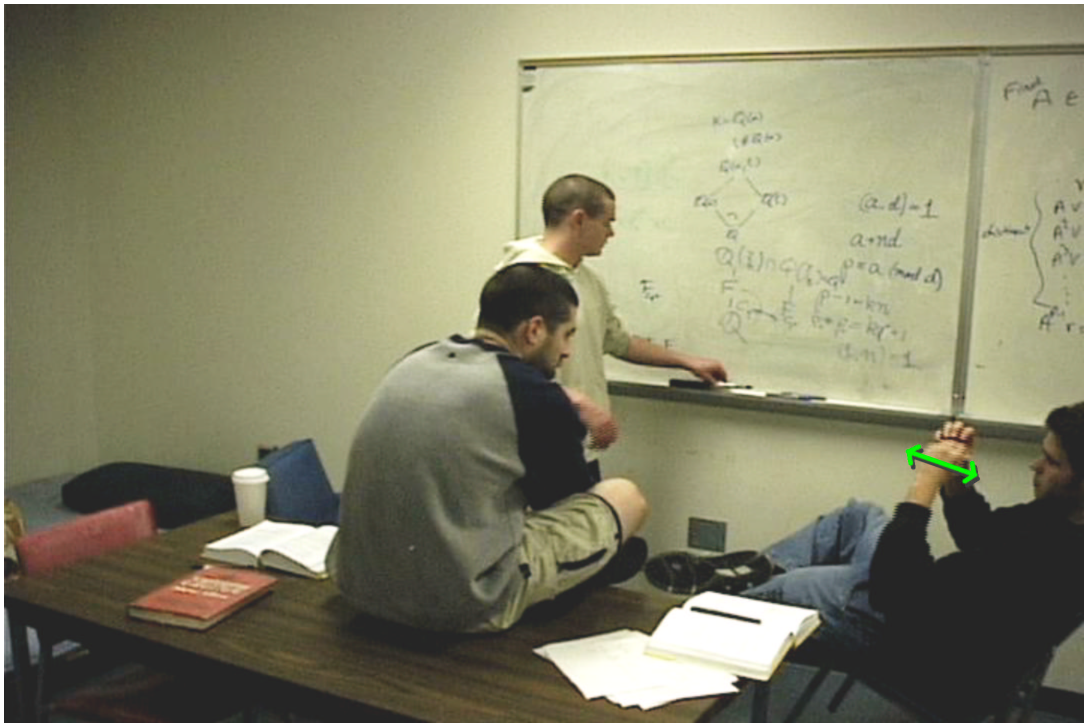

Figure 34: Example (15): but the primitive ones are the ones that like matter

Figure 35: Example (15): then for any zeta

Figure 36: Example (15): to the set of zetas

Figure 37: Example (15): . . . so it's relatively prime to each one of these

Figure 38: Example (15): on this theorem

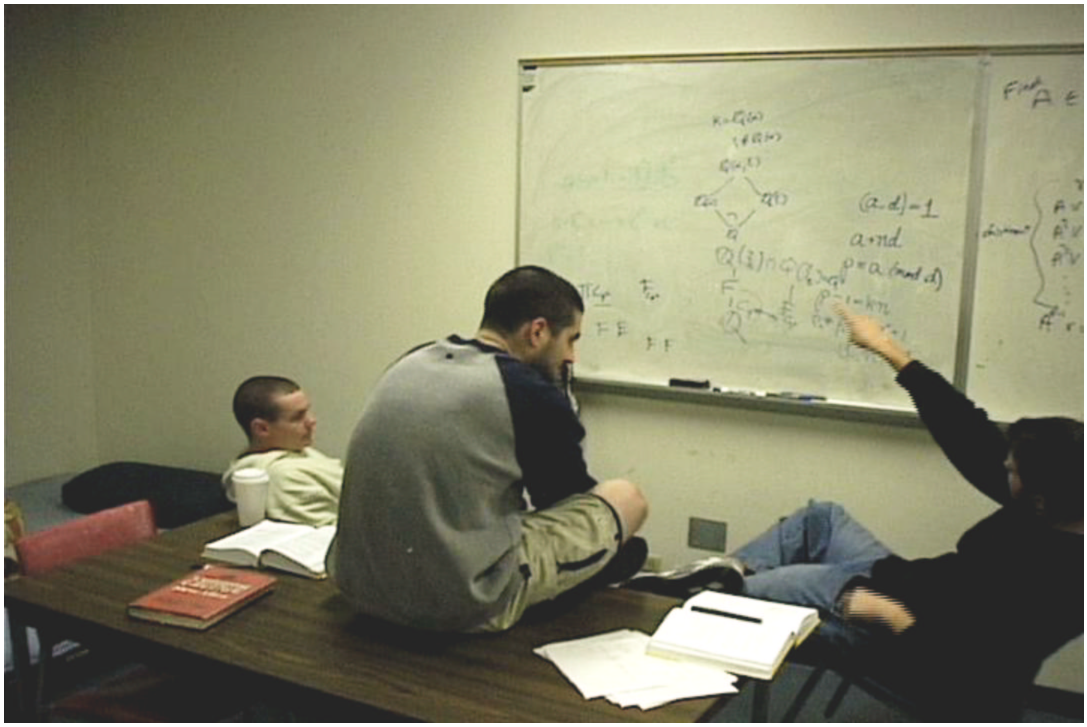

Figure 39: Example (15): that intersection

of him. He maintains this hold while continuing to talk, with overlaid beats when the set is mentioned explicitly. At this point L interrupts to propose a slight modification, rephrasing R's *all the primitive roots of unity* to *all the roots of unity* to communicate the idea that one need not worry about primitive-ness. The accompanying gesture is in a similar place to R's (relative to their respective bodies), and also uses two hands turned inwards to communicate the idea of a set, but he 'rephrases' the space contained to be much larger, with outward motion and wavering in his fingers to code the idea that his set is larger and with less strict attention paid to membership qualifications (metaphorically, this maps to the edges of the set being fuzzy and ill-defined). However, in a way reminiscent of (9), R maintains his hold through L's interruption, and then reaffirms his initial position while reemphasizing his initial gesture.

He then explains his idea further, accompanying it with a series of points — to the

board, with its record of previous discussion, to the space around his hold, referring to the elements of the set, and to the textbook, indicating a part of its contents. Thus in a period of approximately ten seconds, he is using all the forms of anchoring this paper discusses — anchors created in earlier discourse as in (3), a gestural anchor as in (10), and a pre-existing anchor as in (2).

Finally, immediately after this sequence ends, he stands up and pulls out a pen, to start modifying the diagrams on the board while continuing to talk, switching modalities seamlessly.

7 Persistent structure in cognition

Math is hard, and solving math problems is a very difficult task, one that involves very high memory loads, commonly requires holding a complex structure in one's head and then manipulating or reasoning about it, and necessitates tracking complicated chains of inference. We have discussed the way persistent structure is used in discourse to hold topics, maintain complex structures for further elaboration, and organize discussion. There are certain parallelisms between these tasks; one might wonder, then, whether the use of persistent structure might be useful for solving math problems as well as talking about them.

The answer is yes; there is clear evidence in my tapes that students co-opt the linguistic mechanisms for creating real space blends to aid in their general cognition. This seems to be a stronger claim than has previously been made in the literature. Previous commentary on the functions of gesture has observed that gesture increases in cases where lan-

guage is somehow impaired, as when subjects are placed in a delayed auditory feedback condition, or speaking in a foreign language, or aphasic (McNeill 1992; Goldin-Meadow 2003). Furthermore, Goldin-Meadow et al. (2001) show that in a memory task, gesturing while talking causes less distraction than talking alone, suggesting that gesture lightens the cognitive load involved in communication. Others have argued that iconic gesture aids in lexical retrieval, possibly by enhancing memory during lexical search (Rauscher, Krauss, and Chen 1996; Krauss, Chen, and Gottesman 2000), but even this is an argument that gesturing makes talking easier, not an argument that gesturing makes one smarter in general (and it doesn't help that there is also evidence that simple tapping aids in lexical retrieval; see Ravizza 2001). The closest is perhaps Kita (2000) who proposes that one purpose of gesture is to help bring to bear "spatio-motoric thinking", which may be better suited for some tasks than the "analytic thinking" involved in language production. Unfortunately, I am unable from his description to clearly distinguish the two in the context of mathematical reasoning, nor measure the presence of one or the other. My argument will be based instead on the utility of persistent structure.

First, we note that there is reason to believe that mathematical discourse has special need for persistent structure; mathematical jargon has special syntax for just this purpose. Consider some examples:

(16) I ran into *my friend Jack* the other day.

(17) This gives us a set. Take *the pre-image* S' of *this set* S under *an arbitrary function* f.

(18) Let f a function, S a set, M the image of S under f.

(16) is a standard English sentence containing a commaless appositive *friend Jack*. In (17) we see this construction taken to extremes and with slightly different meaning; in

mathese, essentially any NP can be followed by a variable to define an equivalence. Furthermore, there is the special Let construction as in (18) used for defining variables; note that while the copula often appears, it is not mandatory, at least in my dialect. All of these are mechanisms for tying down referents to make sure that they are available for later commentary; mathematicians are rather excessive about this, often naming everything they can whether they will have a need for it later or not[6].

Strangely, though, while these constructions are the bread-and-butter of written or lectured math, they show up rarely in the tapes; when not restricted to writing[7] mathematicians seem to fall back on strategies based on real space blends. One thing about real space blends that appears undescribed to date is that they can be used to scaffold cognition, acting as what have traditionally been called "material anchors" (Hutchins 1995a; Fauconnier and Turner 2002). The theories of real space blends and material anchors are oddly parallel; their definitions are essentially identical, but in the past the two terms have been used by different researchers studying different phenomena — "real space blends" have been used in the analysis of sign language and gesture, and their theory tends to emphasize transient and communicative properties, whereas "material anchors" have been studied in the context of problem solving, with a theory generally emphasizing permanent and designed properties of artifacts[8] I have previously argued that language

[6]In this respect, they are apparently similar to users of American Sign Language, who tend to assign loci to every referent mentioned, whether it will be used again or not. This fact is unsurprising under this analysis. Speakers of verbal languages tend not to have verbal methods to give everything names because gesture usually does a better job at retaining such structure; thus the only people with such methods grammaticized are those who really need them (mathematicians) and those for whom it is no burden (signers).

[7]Lectured math should often be considered to be restricted to writing, since professors must speak to the students's notebooks.

[8]But compare (Hutchins 2002), which discusses the effects of different types of anchors on blend structures.

and gesture can effortlessly put objects traditionally analyzed as material anchors to communicative use; I would now like to suggest that conversely, gestural real space blends can provide cognitive scaffolding for solving problems, i.e., acting as traditional material anchors.

The basic mechanism that makes material anchors useful is offloading, in which one puts various aspects of one's cognitive processing out into the environment to reduce load on the brain (see, e.g., Kirsh 1995; Clark 1997). For example, Kirsh and Maglio (1994) found that expert Tetris players tend to spin the falling shapes around while considering possible places to place them; while they could instead check for possible fits by rotating the pieces mentally, it is faster and takes less effort to let the computer rotate them and simply observe the resulting shape. This is an example of offloading used to solve a purely spatial problem, without a real space blend involved; the trick of making a real space blend is that it allows one to extend this assisted reasoning into abstract domains (Hutchins 1995a; Hutchins 1995b).

Similarly, I postulate that making gestural holds, like writing on paper or whiteboards, can aid in cognition, by using environmental anchoring to reduce memory load, and creating more stable real space blends on which to ring changes. It is rather common to see a group of math students sitting quietly staring at a picture or set of expressions; another case of this sort of thing is shown in example (19).

(19) **R** …[…]…
30 second silent hold, Figure 40

In this example, there is no speech whatsoever. Both women are sitting and writing quietly, absorbed in their own work and entirely ignoring each other. Suddenly the woman on the right pauses, puts her hands up in the air, and stares at them for a full

Figure 40: Example (19)

20–30 seconds. Then she goes back to writing. This is an example of an ordinary gesture, something that one would expect to see in a pure linguistic context, but it is occurring with no communicative purpose at all; clearly the woman finds staring at her hands to be useful in resolving her mathematical difficulty, and knowledge of the problems they work and the common gestural forms they use suggest strongly that her hands are representing some topological space that she needs reason about.

Sometimes, though, the boundary between communicative and cognitive gestures is less clear. Consider (20), in which the woman on the right alternates between explanation and confusion. At first she is describing something about the real line, representing it with the same portion of space as she has used consistently through the whole session, with one hand held as an anchor and the other waving to indicate a region. She then becomes confused, though, leaves her hands up in the air, and stares at them for a time, occasion-

ally moving them to somewhat different locations. When this confusion is resolved, she continues smoothly in the blend, gesturing out the resolution to her difficulty.

(20) **R** it's going to be everything umm…
hands start up in the air, in space used to represent real line
[**after a certain point**
left hand holds still on right side of body, right hand flaps to right of it, Figure 41

L and it's going to be the complement…
to a closed and bounded subset…
***R**'s hands drift to her left, Figure 42*
which will be compact…

R <XXX>
rotates further to the left, still staring blankly at hands, Figure 43

L plus infinity…
so

R what about, okay…
except yeah,
hands shift again, still held still and stared at, Figure 44

L you see how it's the complement to a closed and bounded set

R but, except that, umm

L if it contains the point zero one

R][**oh, yeah, because it has to contain something on the other side—**
hands are released from hold, move around to indicate places on line
on both sides of zero one,
two parallel hands to indicate both sides, Figure 45
and so everything in the middle]
both hands sweep out, scoop, and then back in again, Figure 46

The interesting thing here our inability to answer the question, 'why is she gesturing?' At the beginning it is clearly communicative, and probably is at least partly communicative at the end is as well. The middle portion may also serve a communicative purpose, by informing her interlocutor that she is puzzled; note her continued utterances, that carry little content but indicate that she has not removed herself entirely from the interaction. On the other hand, communication cannot be the whole purpose of the gestures

Figure 41: Example (20): everything after a certain point

Figure 42: Example (20): …

Figure 43: Example (20): <XXX>

Figure 44: Example (20): what about, okay. . . except yeah,. . . but except that, umm

Figure 45: Example (20): on both sides of zero one

Figure 46: Example (20): and so everything in the middle

in the middle portion; there are much simpler and more effective ways to communicate confusion, and we know that staring at a hold like that is a useful way to ground one's visualizations.

Even the final sum-up gesture is ambiguous; it is directed away from her interlocutor, and has the feel of a being self-communicative, using her hands to explain her solution to herself. Our understanding of space is first and foremost a grounded, embodied one; our difficulty is not in applying spatial reasoning to motion and perception, but in avoiding doing so when we wish to reason abstractly. We should therefore consider the physical enactment of a spatial concept to be like understanding it, but more so. Mathematical reasoning is often spatial reasoning dressed up in metaphors and blends (Lakoff and Núñez 2000); it would make sense if gestural enactment enhances mathematical understanding.

Of course, there is no problem with saying that she is gesturing for both communicative and cognitive reasons; there is no incompatibility between them. On the contrary, offloading cognition into communicative channels can be an extremely efficient way to collaborate — one is in a very real sense putting one's thoughts out into a shared space, where they are available to others for inspection. They can even be modified directly if one's interlocutor chooses to manipulate the spatial portion of the space-thought blend. When such blends are shared sufficiently by all parties, it is perhaps more appropriate to consider the entire group as a single cognitive system (cf. Hutchins 1995b; Hutchins 1995a).

8 Conclusion

Humans did not evolve to write mathematical notation while talking. However, it is entirely possible that we evolved to vocalize while engaging in other activity, and coordinating that activity may have been language's original use. As such, it is perhaps not surprising that language use interacts so readily with other actions. In this paper, I have shown many detailed phenomena demonstrating this interaction, in which systematic processes act across language, gesture, and such 'obviously' extralinguistic phenomena as and writing, and drawing. Speech can become writing and back again without disrupting conversational structure; writing and drawing not only look similar to gesture motorically, but follow the same co-timing rules; gesture, drawing, and writing all fulfill similar functions in a similar manner when it comes to creation and maintenance of persistent structure, and what differences we have seen had principled motivation; repetition can be used with speech, gesture, or artifacts to accomplish similar effects, including such basic aspects of online discourse behavior as turn and topic negotiation.

Along the way, I have developed a preliminary theory of *persistent structure*, and used it to explain some of the phenomena listed above. There are a number of mechanisms for maintaining referents in discourse, including various forms of real space blends and repetition, and a number of reasons to do so. Analyzing these together makes clear the connections between them, focuses our attention on which differences are relevant, all the while allowing an analysis that does not attend to modality boundaries — something we have seen is critical in adequately accounting for the data.

Finally, I have demonstrated how strategies developed to maintain persistent struc-

ture in communication are opportunistically co-opted for non-communicative purposes, such as maintaining mathematical concepts in memory while reasoning about them. This draws a connection between two previously disparate lines of research: the sign language and gesture literature on "real space blends" together with the distributed cognition literature on "material anchors" and offloading. That language and gesture should be involved in thought is an old idea, and some particular manifestations have been studied before; for instance, there is a long tradition of boxes labeled "phonological (or verbal, or articulatory) loop" in information-processing models of working memory, e.g. Glanzer and Clark 1963; Baddeley 1986; Burgess and Hitch 1999. This work introduces a new form of this idea by, firstly, observing that material anchors are simply one particular class of real space blend, those anchored by material objects; and secondly, demonstrating that other real space blends, such as those formed in real-time discourse via drawing, writing or gesturing, can also be used for offloading cognition into the world. Moreover, because these techniques are useful simultaneously for communication and cognition, they are uniquely suited for efficient collaboration.

There certainly must be some sorts of boundaries between language, gesture, and everything else; as discussed in §1, language is special in a number of ways. However, these boundaries are lower than generally assumed, and we do not yet know their location. Modularism (Fodor 1983) has many flaws, but it does at least focus attention on the question of interfaces. Today we know that interfaces cannot be so slim and neat as modularists would hope, and this work only continues that trend. Nevertheless, we know from neural architecture that there are articulated subsystems in the brain, and that connectivity between subsystems is far lower than within them; it is impossible that there

be no interfaces whatsoever, that everything in the brain has full access to the internal structure of everything else.

Empirical research is needed to determine the nature of these boundaries, and the results will have important implications for theories of language and mind. The phenomena demonstrated in this paper are not mere minutiae; they are radically incompatible with many serious theories of language and cognition. More importantly, the present results are not simply negative; while they falsify some ideas, they simultaneously point towards new theories that better match the empirical results. Overall this work is quite preliminary in nature, focusing on proving possibility and speculating on mechanism; but it begins to elucidate some important aspects of the relation between language and mind, and holds the potential for new approaches that may change our conception of both.

Acknowledgments

This thesis has been long in the making, and as the culmination of my undergraduate career at UC Berkeley, has been influenced by many people. To keep the list finite, I'll be brief; I would like to thank and acknowledge Eve Sweetser and the rest of the Berkeley Gesture Project (and of course Fey Parrill, who started it all), a number of anonymous subjects, Dan Slobin, George Lakoff, Jerry Feldman and the rest of the Neural Theory of Language project, Rafael Núñez, Teenie Matlock, Monica Gonzalez-Marquez, Irene Mittelberg, Zack Weinberg, my parents, as well as many others who have encouraged me, inspired me, and otherwise helped me along my path.

And, of course, Shweta Narayan, whose forebearance has been remarkable, advice and assistance indispensable, and without whom.

Of course, any errors or flaws remaining are my own fault for being too mule-headed to take their good advice.

References

Baddeley, A. (1986). *Working memory*. New York: Clarendon Press/Oxford University Press.

Burgess, N. and G. J. Hitch (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review 106*(3), 551–581.

Butcher, C. and S. Goldin-Meadow (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come together. See McNeill (2000b), pp. 235–257.

Clark, A. (1997). *Being there: Putting body, brain, and world together again*. Cambridge, MA: MIT Press.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. (2003). Pointing and placing. See Kita (2003), pp. 243–268.

Clark, H. H. and J. E. Fox Tree (2002). Using *uh* and *um* in spontaneous speaking. *Cognition 84*, 73–111.

Coulson, S. (2001). *Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction*. Cambridge: Cambridge University Press.

Drummond, K. and R. Hopper (1993). Back channels revisited: acknowledgement tokens and speakership incipiency. *Research on Language and Social Interaction 26*(2), 157–177.

Fauconnier, G. (1985). *Mental spaces*. Cambridge, MA: MIT Press.

Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge: Cambridge University Press.

Fauconnier, G. and M. Turner (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York, New York: Basic Books.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Furuyama, N. (2000). Gestural interaction between the instructer and the learner in *origami* instruction. See McNeill (2000b), pp. 99–117.

Glanzer, M. and W. H. Clark (1963). Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning & Verbal Behavior 1*(4), 289–299.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: The Belknap Press of Harvard University Press.

Goldin-Meadow, S., H. Nusbaum, S. D. Kelly, and S. Wagner (2001). Explaining math: Gesture lightens the load. *Psychological Science 12*(6), 516–522.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics 32*(10), 1489–1522.

Goodwin, C. (2003). Pointing as situated practice. See Kita (2003), pp. 217–242.

Goodwin, M. H. and C. Goodwin (1987). Children's arguing. In S. U. Philips, S. Steele, and C. Tanz (Eds.), *Language, Gender, and Sex in Comparative Perspective*, pp. 200–248. Cambridge: Cambridge University Press.

Halliday, M. A. K. and R. Hasan (1976). *Cohesion in English*. London: Longman.

Hanks, W. F. (1990). *Referential practice: Language and lived space among the Maya*. Chicago, IL: University of Chicago Press.

Haviland, J. B. (1993). Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology 3*, 3–45.

Haviland, J. B. (2000). Pointing, gesture spaces, and mental maps. See McNeill (2000b), pp. 13–46.

Hindmarsh, J. and C. Heath (2000). Embodied reference: A study of deixis in workplace interaction. *Journal of Pragmatics 32*(12), 1855–1878.

Hoque, S. (2003). Gesture in architectural discourse: Meaning and relevance. Master's thesis, Architecture Department, University of California at Berkeley.

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge, MA: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science 19*(3), 265–288.

Hutchins, E. (2002, August). Material anchors for conceptual blends. In *Proceedings of The Way We Think Symposium on Conceptual Blending*, Number 23 in Odense Working Papers in Language and Communication, Odense, Denmark, pp. 85–111. University of Southern Denmark.

Hutchins, E. and L. Palen (1997). Constructing meaning from space, gesture, and speech. In C. P. Lauren B. Resnick, Roger Saljo and B. Burge (Eds.), *Discourse, Tools, and Reasoning: Essays on Situated Cognition*, pp. 23–40. Germany: Springer-Verlag.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence 73*(1–2), 31–68.

Kirsh, D. and P. P. Maglio (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science 18*(4), 513–549.

Kita, S. (2000). How representational gestures help speaking. See McNeill (2000b), pp. 162–185.

Kita, S. (Ed.) (2003). *Pointing: Where Language, Culture, and Cognition Meet*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Krauss, R. M., Y. Chen, and R. F. Gottesman (2000). Lexical gestures and lexical access: A process model. See McNeill (2000b), pp. 261–283.

Lakoff, G. (1996). Sorry, I'm not myself today: The metaphor system for conceptualizing the self. In G. Fauconnier and E. Sweetser (Eds.), *Spaces, Worlds, and Grammar*. Chicago: University of Chicago Press.

Lakoff, G. and R. Núñez (2000). *Where Mathematics Comes From*. New York, New York: Basic Books.

LeBaron, C. and J. Streeck (2000). Gestures, knowledge, and the world. See McNeill (2000b), pp. 118–138.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.

Mayberry, R. I. and J. Jaques (2000). Gesture production during stuttered speech: insights into the nature of gesture-speech integration. See McNeill (2000b), pp. 199–214.

McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2000a). Catchments and contexts: non-modular factors in speech and gesture production. See McNeill (2000b), pp. 312–328.

McNeill, D. (Ed.) (2000b). *Language and gesture*. Cambridge: Cambridge University Press.

McNeill, D., F. Quek, K.-E. McCullough, S. Duncan, N. Furuyama, R. Bryll, X.-F. Ma, and R. Ansari (2001). Catchments, prosody and discourse. *Gesture 1*(1), 9–33.

Ochs, E., P. Gonzales, and S. Jacoby (1996). "when I come down I'm in the domain state": grammar and graphic representation in the interpretive activity of physicists. In E. Ochs, E. A. Schegloff, and S. A. Thompson (Eds.), *Interaction and grammar*, pp. 328–369. Cambridge: Cambridge University Press.

Ochs, E., S. Jacoby, and P. Gonzales (1994). Interpretive journeys: How physicists talk and travel through graphic space. *Configurations* (1), 151–171.

Özyürek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. See McNeill (2000b), pp. 64–83.

Rauscher, F. H., R. M. Krauss, and Y. Chen (1996, July). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science 7*(4), 226–231.

Ravizza, S. M. (2001). *Effects of movement on lexical retrieval*. Ph. D. thesis, University of California at Berkeley.

Tabensky, A. (2001). Gesture and speech rephrasings in conversation. *Gesture 1*(2), 213–235.

Tannen, D. (1989). *Talking Voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge: Cambridge University Press.

Wilkins, D. P. (1997). Alternative representations of space: Arrernte narratives in sand and sign. In M. Biemans and J. van de Weijer (Eds.), *Proceedings of the CLS Opening Academic Year '97–'98*, Tillburg, pp. 133–162. Center for Language Studies.

Yngve, V. H. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting, Chicago Linguistic Society*, Chicago, IL, pp. 567–578. Chicago Linguistic Society.