UNIVERSITY OF CALIFORNIA, SAN DIEGO

Reference in action: Links between pointing and language

A dissertation submitted in partial satisfaction of the requirement for the degree Doctor of Philosophy

in

Cognitive Science

by

Kensy Andrew Cooperrider

Committee in charge:

Professor Rafael Núñez, Chair
Professor Seana Coulson
Professor Karen Emmorey
Professor John B. Haviland
Professor Edwin Hutchins

2011

The Dissertation of Kensy Andrew Cooperrider is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2011

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

An outsized thanks goes, first, to my academic mentors at UCSD. This list includes my advisor, Rafael Núñez, as well as the other members of my dissertation committee: Seana Coulson, Karen Emmorey, John B. Haviland, and Edwin Hutchins. Together they have taught me— or tried, anyway, to teach me— too many lessons to enumerate here, but chief among these are: how to design research, how to argue from a place of critical sympathy, how to stay focused on big questions, and, not least, how to have fun in academia. Their influence runs throughout these pages. Other mentors— especially Marta Kutas and Benjamin Bergen— have offered informal guidance and vigorous criticism for which I am grateful. I am confident that many of my mentors' best lessons are yet to be learned, as I carry with me the formidable examples they have provided.

Much of the research reported in this dissertation could not have been done without the efforts of undergraduate research assistants in the Embodied Cognition Lab. I am grateful for their diligence, unassuming criticism, and patience throughout countless false starts. In particular, I wish to thank Emma Bloom, Richard Chen, and Rachel Cusing.

Many different intellectual communities nurtured my work at different points during my UCSD career. These include: Rafael Núñez's Embodied Cognition Lab, Seana Coulson's Brain and Cognition Lab, Gesture Group San Diego, the Center for Research in Language, John B. Haviland's Linguistic Anthropology Lab, and, for a brief stint, the Gesture Group at UC–Berkeley, led by Eve Sweetser. I am very grateful to the members of these communities for feedback, insight, and camaraderie. One nurturing micro-

community deserves special mention and that is my entering cohort. The first three years of graduate school are not without travail, and your support was pretty epic.

The data presented in Chapter 4 was collected during a collaborative field trip to Papua New Guinea in August–September of 2009. For warmly welcoming me along on the trip, and for companionship in the field, I would like to thank Rafael Núñez, Jürg Wassmann, and Verena Keck. I never would have been able to participate in this research were it not for a fellowship awarded in the spring of 2009 by the Friends of the International Center for which I am extremely grateful. After returning from the field, James Slotta was generous in answering many ham-fistedly posed questions about the Yupno language.

Other funding was provided by a UC pre-doctoral fellowship in the humanities, a pre-doctoral traineeship from the CRL, and a Glushko award. Travel funding was provided by the Dean of Social Sciences and the Cognitive Science department. All this financial support is most gratefully acknowledged.

Thanks to my friends who made San Diego a home— and a festive one at that— for their support. The Earthpigs, the Smalls crew, devotees of the Cog Sci happy hours, and Hillcrest shuttlers— many thanks for nourishments of different kinds. Exuberant thanks to Jordan Davison, my closest fellow-traveler throughout the dissertation-writing process, for abiding much Fribble-related hand-wringing and unsolicited nose-pointing. And, finally, thanks to a much larger network of colleagues and friends, at UCSD and far beyond, who in recent years have unwittingly sustained me with tiny encouragements.

I gratefully acknowledge the permission of my co-author, Rafael Núñez, to include Chapter 4, ‘Nose-pointing in Yupno’, in this dissertation.

VITA

1999 - 2003	B.A. in Linguistics Duke University, Durham, NC
2005 - 2011	Ph.D. in Cognitive Science University of California, San Diego, La Jolla, CA

ABSTRACT OF THE DISSERTATION

Reference in action: Links between pointing and language

by

Kensy Andrew Cooperrider

Doctor of Philosophy in Cognitive Science

University of California, San Diego, 2011

Professor Rafael Núñez, Chair

When referring to things in the world, speakers produce utterances that are composites of speech and action. Pointing gestures are a pervasive part of such composite utterances, but many questions remain about exactly how pointing is integrated with speech. In this dissertation I present three strands of research that investigate relations of different kinds between pointing and language.

A first strand investigates the relationship between pointing gestures and spoken demonstratives, such as *this* and *that* in English. Linguists, philosophers, and psychologists have long noted the pointing-demonstrative relationship but have not yet

characterized it with any precision. At the same time, cross-disciplinary controversy about the meaning of demonstratives has intensified. I present findings from two studies using a referential communication task, which suggest that demonstrative use may hinge on pointing in previously unappreciated ways.

A second strand of research presents an exploratory investigation of a commonplace but ignored class of pointing: gestures that speakers direct toward their own bodies. An analysis was carried out of body-directed gestures in a corpus of one-on-one interviews, resulting in a typology of three types of body-directed gestures— *self-points*, *body-points*, and *body-anchors*. Each type is considered in turn, with a focus on basic questions about how, when, and why such gestures are produced.

The third strand considers cross-cultural differences in pointing and the question of what motivates these differences. A case study is presented of a previously undocumented facial pointing gesture— *nose-pointing*— used by the Yupno, an indigenous group of Papua New Guinea. Based on examples of pointing and non-pointing uses of the form, we propose that facial gesture is linked to a particular semantic theme, and discuss how this link is both iconically motivated and shaped by features of Yupno language and communicative practice.

Together these different strands of research contribute to our understanding of pointing as both a window into processes of conceptualization and a cornerstone of human social interaction. The findings presented offer new insights into the disparate forces— biomechanics, grammar, conceptual structure, and cultural practices— that give this cornerstone shape.

Chapter 1. Introduction

1.1. A pointing scene

On April 24, 1066 AD, Halley's Comet blazed across the night sky. The Bayeux Tapestry, which was constructed just years later in England, records the event, depicting the comet as a toothy ball of fire. The tapestry also renders the human response. Beneath the words ISTI MIRANT STELLA (or, THESE PEOPLE WONDER AT A STAR), six men cluster on the ramparts of a castle, their heads tilted skyward (see Figure 1.1). Five of the men are pointing, index fingers extending up toward the fiery orb. One of the men, standing a little apart from the others, turns back to his fellow onlookers as if to confirm that they are sharing his experience. It is an immediately recognizable human scene: frozen in poses of attention, the embroidered figures embody a familiar— even elemental— social scenario, a scenario marked by shared orientation to the natural world and to each other. It is a tableau of human social interaction, centered on the pointing gesture.

When we think of pointing gestures, we perhaps imagine something very like this tableau. In our folk theories, the pointing gesture is often woven into scenarios in which one person deliberately orients another to some feature of the world. Perhaps we imagine, also, the pointer's extended index finger and, subsequently, the audience's shifted gaze. But this tableau— and the folk view that it seems to capture— is in certain ways misleading. A central aim of this dissertation is to give this scene life, to ground it in empirical observation, to explore pointing gestures in action, and, in particular, to investigate in detail a crucial dimension of pointing that is entirely absent from the tapestry's rendering of the comet scene: its partnership with spoken language.

Figure 1.1: A scene from the Bayeux Tapestry, which dates from the late 11th century AD, depicting the response to the appearance of Halley's comet. Image credit: Boris Doesborg.

Pointing is a hallmark of human communication. Starting sometime before we can talk and continuing throughout our communicative lives, humans point to other objects, to places, and to other people; we point in all kinds of settings, for all kinds of purposes. The pointing gesture has been described as a basic building block of human interaction (Kita, 2003), and it is widely considered both universal to— and distinctive of— our species (Liszkowski & Tomasello, 2011). Given the ubiquity of pointing across cultures, across contexts, and across the lifespan— not to mention its almost mythical status as a

communicative primitive— it is somewhat surprising to confront the scarcity of research on when, how, and why people point. Empirical research on pointing to date has been largely concentrated in a handful of areas, such as the relationship between pointing and language acquisition in infants (e.g. Bates & Dick, 2002), and the question of whether nonhuman animals produce (e.g. Veà & Sabater-Pi, 1998; Povinelli et al., 2003) or comprehend pointing gestures (e.g. Hare & Tomasello, 2005). Another stream of work— considerably more speculative but with a long history— stitches pointing centrally into grand narratives about the very origins of humanness (Thao, 1984 [orig. 1973]; Tomasello, 2008). Much of this work treats pointing as a monolith, a capacity one either has or does not. Considerably rarer have been investigations of pointing that delve into its particulars and that seek to explain it as a patterned behavior, richly and curiously entangled with speech.

In fact, even within gestures studies proper, interest in pointing has been to a considerable extent eclipsed by interest in iconic gestures. This has been especially the case among researchers with an interest in cognitive dimensions of gesture. Classic tasks used by psychologists to elicit "gesture" are often only good at eliciting *iconic* gestures. The 'Canary Row' cartoon used by McNeill and colleagues, for instance, prompts vanishingly low rates of pointing (McNeill, 1992, pg. 285). Production models of gesture often only aim— without apology— to explain imagistic gestures (e.g. Hostetter & Alibali, 2008). Even when pointing occurs in abundance in psychological studies, it is occasionally just lumped in with other gestures (e.g. Wagner et al., 2004). What factors might explain such systematic neglect? Fauconnier (1997, pg. 32), in a different context, has discussed two illusions that serve to direct scholarly attention away from certain

phenomena in the social sciences: the *illusion of simplicity* and the *illusion of rarity*. Both of these illusions appear to be at play in the case of pointing. Pointing is often reduced to a particular prototype of pointing, complete with a characteristic morphology— the extended index finger— and embedded within a prototypical interactive scenario[1]. At the same time that it is caricatured as simple, pointing is often regarded as rare, perhaps confined to cases of direction-giving or of ostension-based instruction. Under the spell of these twin illusions, basic questions about pointing have gone under-explored. A central but often overlooked aspect of pointing, for instance, is the nature of its links with speech. How do pointing gestures and speech combine? At what levels of analysis do we see such links and with what degrees of systematicity? The question turns out to be one of considerable complexity, and it is this complexity that is the focus of the present dissertation.

1.2. How gesture and speech relate: From co-expressivity to co-patterning

Gesture and speech are often said to be both *co-timed* and *co-expressive*. What is meant by the latter term is that the meaning of the spoken parts of an utterance *align* in some sense with the meaning of its gestural parts. Co-expressivity is one of the bedrock observations around which interest in gesture has recently flourished; its exact nature and cognitive implications are of ongoing theoretical concern in all corners of gesture studies. Several formulations of the notion of co-expressivity have been put forward in recent years, with only minor differences in emphasis. Adam Kendon, for example, has written

[1] Lakoff (1987, pg. 490-1) and more recently Langacker (2008, pg. 284) have both discussed the idea that the category of "pointing" is structured around a particular prototype.

influentially on gesture and speech as two aspects of the same process of utterance production. He writes that "[t]he relationship between word and gesture is a reciprocal one— the gestural component and the spoken component interact with one another to create a precise and vivid understanding" (Kendon, 2004, pg. 174). David McNeill has offered a related formulation: "[G]esture and speech express the same underlying idea unit but express it in their own ways— their own aspects of it, and when they express overlapping aspects they do so in distinctive ways" (McNeill, 2005, pg. 25). Part of what makes the contributions of gestural and spoken components distinctive is that, as McNeill has repeatedly emphasized, gesture and speech have fundamentally different properties: gesture is idiosyncratic and analogue, while speech is conventionalized and discrete. In a strict sense, then, gestural and spoken components cannot be "redundant", "over-lapping", or "matching" (c.f. McNeill, 2000, pg. 6). What we find instead is that speech-gesture combinations exhibit complementarity of different sorts, the modalities serving to "mutually elaborate" each other (Goodwin, 1995, pg. 613).

The focus of the above treatments is invariably the relationship between gesture and its immediately co-occuring speech, most often at the level of the word. The relation between a gesture and its "lexical affiliate" (Schegloff, 1984)— as the immediately co-produced word is sometimes called— is without a doubt a conspicuous and theoretically interesting kind of co-expressivity. But it is not the only kind. In the present dissertation, the notion of co-expressivity is opened out substantially. It is recast to encompass the *co-patterning* of aspects of spoken discourse at *any level of analysis* with aspects of gesture at *any level of analysis*. Levels of analysis in the speech stream that may be implicated in such co-patterning are many, and include the prosodic level, the morphological level, the

lexical level, the information-structural level, the level of sociolinguistic register, and so on. Levels of analysis in the gestural stream may include the very presence or absence of gesture, as well as aspects of the gestural morphology such as the handshape, movement dynamics, and what I will be calling robustness.

Speech-gesture co-patterning as just described is a broad notion indeed, and it is a substantial empirical task to demonstrate empirically what kinds of links are observed and what kinds are not. In the case of pointing gestures, much of this empirical work still lies ahead. This is because speech-pointing relations are often presumed innocent— that is, unmarked by cognitive or semiotic complexity. In the next section, after definitional preliminaries, I very briefly sketch three specific issues in the area of pointing-speech co-patterning. These issues are: first, the putatively privileged relationship between pointing gestures and deictic words; second, the issue of variation in pointing morphology and its possible relations to speech; and, third, the issue of pointing and its lexical affiliates.

1.3. Pointing and language

1.3.1. Definitions of pointing

Typologies of gesture have invariably included a category for pointing. Wilhelm Wundt distinguished *demonstratives* from *descriptives*; David Efron drew a line between *deictics* and *physiographics*; Ekman and Friesen set *deictic movements* apart from five other kinds of illustrators; and, perhaps most influentially for contemporary researchers, David McNeill has drawn a five-way distinction between *deictics*, *iconics*, *metaphorics*, *beats*, and *emblems* (Kendon, 2004). What is it that makes pointing special, if indeed it is? On what grounds— morphological, functional, or otherwise— is it distinguished from

other kinds of gestures? Though in the West pointing is strongly identified with a particular prototype with certain features, as mentioned above, there is a much wider world of pointing that deserves empirical attention. The category is fuzzily bordered, but at its kernel is the idea that pointing involves a "movement toward" some region in space (Eco, 1976, pg. 119; discussed in Kendon, 2004, pg. 200), and that this movement carries attention-directing intent. Exactly what is moved toward, which articulators are used to execute the movement, and with what degree of intensity, however, are all subject to considerable variation. One clear point of contrast between pointing gestures and other gestures is in how they are treated by the audience. While all gestures cry out for attention, pointing gestures immediately deflect that attention elsewhere.

Complicating definitional matters somewhat is the fact that pointing gestures are not always pure. They often hybridize with iconic elements. McNeill (2005) has recently stressed that gestures are best understood not as belonging to discrete *types*, but as exhibiting different *dimensions*— deictic, iconic, metaphoric, and beat— to various degrees. Kendon (2004) seems to largely share this view, as evident in the following caveat about the identification of pointing gestures: "In general, gestures may be said to vary in the degree to which they show a deictic component. Gestures that are said to be pointing gestures are dominated by the deictic component almost to the exclusion of everything else. We may say of such gestures that they are *specialized* as pointing gestures" (2004, pg. 205, orig. italics). While on the one hand it bears mention that gestures are always semiotically composite (c.f. Enfield, 2009)— that is, blends of indexical, iconic, and symbolic features— on the other hand, it will be useful as an expository shortcut to retain the label "pointing" for those gestures whose most

pronounced semiotic dimension appears to be indexical.

1.3.2. Pointing gestures and pointing words

Numerous commentators, from descriptive linguists to philosophers, have remarked on an apparently special association between pointing gestures and deictic words[2]. For instance, Hanks (2005) writes: "[W]hile both indexicality and gesture are pervasive in language, referential deictics are unique in joining the two systematically" (pg. 195). Yet Hanks himself provides no account of the nature of this systematicity, focusing instead on verbal deixis and the circumstances shaping its use. A more specific claim often advanced is that there is a privileged relation between pointing and a subset of deictic words, *demonstratives*. (Other deictic words such as *I* and *now* are not assumed to require gestural supplement in the same way.) In English this category includes both entity-referring demonstratives, *this*, *that, these,* and *those,* and place-referring demonstratives, *here* and *there*, as well as phrasal manner demonstratives, *like this* and *like that*. Diessel (2006) makes much of the close relationship between pointing gesture and demonstratives, noting that they apparently serve "the same function" (pg. 470) and commonly co-occur in adult speech. In support of these claims, however, he largely cites earlier theoretical treatments making similarly in-passing assertions, not empirically based reports. In general, the link between pointing and demonstratives has a taken-for-granted quality, and very few have endeavored to explore the link in a rigorous,

---

[2] For earlier sources of this idea, see Frege (1967 [orig. 1918]), whose discussion of verbal deixis includes the parenthetical comment that "[t]he pointing of fingers, hand movements, glances may belong here too" (pg. 24). See also extensive discussion in Buhler (1990 [orig. 1934]).

empirically grounded way. It is evidently not the case that all demonstrative words involve a pointing gesture; nor is it that case that all pointing gestures are co-produced with a demonstrative. What more can be said of their relation in light of this clear dissociation?

More specific ideas about the relationship between demonstratives and pointing have occasionally been entertained. Levelt et al. (1985), for example, argue that for certain deictic words gestural supplementation is obligatory:

> Deictic terms, such as *here*, *there*, *I*, *you*, *this*, *that*, derive their interpretation in part from the speaker/ listener situation in which the utterance is made. Among these terms only *here*, *I*, and in some cases *you* are directly referential; given the situation their reference is unambiguous. The other deictic terms, however, require the speaker to make some form of pointing gesture, for example, by nodding the head, visibly directing the gaze, turning the body, or moving arm and hand in the appropriate direction (pg. 134).

While it may be an overstatement to say that gestures are strictly obligatory in such cases, it is certainly plausible that different deictic words differ in the nature of their association with pointing, and that these differences are more or less stable across contexts. Bangerter (2004) explored pointing-demonstrative relations in the context of a controlled referential communication task in which speakers identified human faces from arrays placed at different distances. He found that speakers only use pointing and demonstratives in combination when pointing is effective. Put differently, as speakers become less able to point out targets unambiguously, their use of pointing-demonstrative combinations drops off markedly. (Interestingly, however, speakers continue to point at high rates at all distances from the array, but without demonstratives words.) The choice of whether or not to point thus clearly interacts with the choice of whether or not to use a

demonstrative; their partnership appears to be very much the result of design. Along similar lines, Kendon (2004) notes, somewhat impressionistically, that there is a relationship between deictic words and pointing morphology. He writes: "It is particularly interesting to note that whenever the index finger is used in pointing, the speaker also often employs a deictic word, whereas when the Open Hand is used, deictic words in the associated speech are less often observed" (pg. 208). Of considerable further interest is the question of whether pointing might interact in some way with the choice of *which* demonstrative form (*distal/ proximal*, *singular/ plural*) to use. For example, does the ability to point unambiguously affect the choice of demonstrative form? Are different demonstrative forms differentially associated with pointing? These questions are taken up in considerable detail in Chapter 2, 'Pointing and the meaning of spoken demonstratives'. In sum, while there is broad agreement that pointing gestures and pointing words are linked both developmentally and synchronically, much remains to be understood about the particulars of this relationship.

1.3.3. Morphological variety in pointing

A key issue in understanding the co-patterning of speech and pointing is the question of whether—and how— differences in pointing form relate to differences in the accompanying speech. In other words, what regularities are observed in how different pointing morphologies are paired with speech? And how systematic are any such regularities? If morphological variety is shown to be largely unconstrained and chaotic, the theoretical interest of the phenomenon may be minimal. If, on the other hand, it is shown to co-pattern with speech, the theoretical interest becomes considerable: it

suggests a linking up of gesture and speech at a previously unappreciated granularity, a linkage with important implications for the cognitive processes involved in speech production. Formal variety in pointing gestures occurs along a number of different parameters. Only three will be briefly considered below: the handshape used, the overall intensity of the pointing action, and the articulator used.

A first parameter of morphological variation is the handshape used. Extension of the index finger is certainly a prototypical feature of pointing in the West, but it is neither a necessary nor sufficient feature of pointing. Considerable variation in pointing handshape has been noted by several authors. Kendon (2004) describes handshape and palm orientation variants in use by both Italian and British English speakers, and links these variants to differences in "the way the object being referred to is presented in the speaker's discourse" (pg. 201). For instance, pointing with the palm oriented upwards is associated with a theme of "presenting". It is not entirely clear from Kendon's discussion whether these differences are considered to be strictly conventional— and thus expectably variable across cultures— or instead *motivated* in some way— and thus expectably robust across cultures. The interplay of motivated-ness and conventionality in handshape variation will be taken up in Chapter 3, 'Body-directed gestures'. Wilkins (2003) has described the use of a set of highly conventionalized handshape-meaning pairings in Arrernte, an Australian aboriginal language, such as the use of a flat hand, palm oriented vertically, when indicating cardinally oriented paths. Intriguingly, he also notes more subtle differences between Arrernte and Western pointing, such as the fact that, even when pointing with the index finger extended, Arrernte speakers do not necessarily bunch their remaining fingers into the palm in the same way that Westerners

do. Unfortunately, at present, there are no studies describing regularities in pointing handshapes among speakers of American English.

A second parameter of morphological variation in pointing is the degree of intensity, or *robustness*, with which the gesture is produced. Robustness is reflected in a suite of features: the degree of arm extension, the presence or absence of co-oriented gaze, and certain aspects of the dynamics of the movement, such as the duration of the apical hold. Does robustness co-pattern with features of the co-produced speech? Drawing on data from locality interviews in Lao, Enfield et al. (2007) have demonstrated that the robustness of a pointing gesture is relatable to discourse structure: more robust gestures (in the authors' terminology "B-points" for "big") are characteristic of location-focus utterances, while less robust gestures ("S-points" for "small") are associated with non-focal reference to location[3]. The finding constitutes an important conceptual proof of the idea that pointing co-patterns with discourse structure, and thus opens the door to a number of further questions. One such question, which will be taken up in some detail in Chapter 3, is whether discourse structure conditions— not the form of pointing *per se*— but the gesture's very presence or absence.

Another aspect of pointing morphology is the choice of articulator used. Again, when we look beyond the canonical prototype we find considerable variation, both within and across cultures, in which articulators are used. Even among English speakers, it has been observed that the head is used deictically (McClave, 2000), and that other

[3] I prefer the term robustness over Enfield et al.'s (2007) size-based monikers— "B-points" and "S-points"— which would appear to privilege the degree of arm extension. As Enfield et al. note, the involvement of gaze and torso orientation are also important, and it is an empirical question whether any of these parameters has a privileged relationship to information structure.

articulators— an elbow, a foot, or a pool cue— may be conscripted on an *ad hoc* basis. No hypotheses have been ventured about why English speakers choose non-manual articulators, though it is perhaps presumed to be driven by unavailability of the hands. Looking across cultures, it is now well documented that some groups make more extensive use of non-manual pointing than do English speakers. In particular, *lip-pointing*— a protrusion of the lips in concert with a lifting of the head— has been documented in detail among the Cuna of Panama (Sherzer, 1972) and among Lao speakers (Enfield, 2001), and it is now reported to be used quite broadly (see Wilkins, 2003, pgs. 175-9 for discussion). Importantly, where lip-pointing is present it is found to co-exist with finger-pointing, not replace it altogether. The question thus arises of whether— and how— lip-pointing contrasts with manual pointing. Enfield (2001) argues that lip-pointing among Lao speakers carries a *recognitional* shading, that it signals a "you know the one I mean" attitude by the speaker. Lip-pointing thus differs in important pragmatic respects from other available pointing options in the Lao system. Chapter 4, 'Nose-pointing in Yupno', introduces another culturally specific type of non-manual pointing, and one which contrasts in several key respects with lip-pointing.

1.3.4. Lexical affiliates of pointing gestures

What kinds of relations are observed between pointing gestures and their immediately co-produced speech? What is the relation between what is pointed to and what is concurrently said? Very little empirical work has addressed this question directly, perhaps on the assumption that such relationships are transparent. It is assumed, perhaps, that what is said refers to whatever is pointed to in a straightforward way: when pointing

to a man one might say “that guy”, when pointing to a meatloaf “the meatloaf”, and so on. As Clark et al. (1983) and others have noted, however, this assumption is empirically false: what is pointed to— in their terms, the *demonstratum*— is not always the same thing as the what is referred to concurrently in speech— in their terms the *indicatum* (see also Miller, 1982). They offer a hypothetical example in which a speaker points to a newspaper and says "I used to work for those people" to mean that she was previously employed by the New York Times company. The physical copy of the newspaper and the intended referent thus stand in a metonymic relation. In more recent work, Clark (2003) has proposed the notion of a *chain of indicating* to characterize such cases. Echoing this formulation, Streeck (2009) refers without elaboration to similar cases as instances of “indexical reference once removed” (pg. 143).

An ethnographically well-attested instance of this kind of complexity occurs in the course of referring to persons (see Haviland, 1993; Stivers & Enfield, 2006). A speaker, upon first mention of someone, points toward a place associated with that person. The interpretability of the reference does not depend on whether the person is actually at the indicated place at the time of utterance; habitual association is enough. Levinson (2006, pgs. 58-61) describes a particularly striking example, in which a man attempts to achieve reference to a young woman. He twice refers to her verbally as "that girl" and both times accompanies the phrase with a pointing gesture— but the gestures point in different directions. The man first points "over the mountain to... where the girl was raised, and then West where she has just died" (pg. 60). Haviland (2000, pg. 32) discusses a similarly striking case. There is little reason to suppose that referential complexity of this sort is confined to cases of reference to third parties. Indeed, research

presented in Chapter 3 describes another rich empirical arena in which to explore the variety of relations between pointing and its lexical affiliates: gestures directed at the speaker's own body.

1.4. Overview of chapters

This dissertation's three empirical chapters aim to illuminate the issues just outlined. Owing to the vastness of the unexplored terrain, however, coverage is necessarily more scattershot than programmatic. Each chapter uses a different method and zooms in on a particular empirical arena, providing a snapshot of co-patterning in pointing and speech.

Chapter 2, 'Pointing and the meaning of spoken demonstratives', examines the relationship between pointing gestures and English demonstratives— *this, that, here, there*. The chief aim is to contribute to a long-standing and recently rekindled debate about the semantics of demonstrative words by considering more closely multimodal aspects of their production. Linguists have long analyzed demonstratives as encoding a distance contrast— proximal *this*/ distal *that*— from the speaker, but several theorists have more recently argued that non-spatial contextual parameters are primary. To date, however, large corpora of comparable instances of naturalistic demonstrative use have not been brought to bear on this debate. Two controlled naturalistic studies are presented that test the long-standing assumption that demonstrative use is grounded in *distance contrasts* against the alternative possibility that demonstrative use is grounded in *pointing affordances*. More specifically, the alternative hypothesis is that the choice of proximal over distal demonstrative forms in referring to entities is motivated— at least in part— by

the speaker's ability to successfully direct attention to those entities. According to this hypothesis, speakers favor proximal forms when they can point unambiguously, and favor distal forms or non-demonstrative referring expressions when they cannot. In the present studies participants worked together, using both speech and gesture, to identify alien creatures from projected arrays. Each demonstrative reference was analyzed for a number of features: which alien target it referred to, the target's distance, and whether the demonstrative was accompanied by pointing. In addition to evaluating the particular hypothesis about effects of pointing affordances on demonstrative use, the study aimed to generate a substantial corpus in which other patterns in the partnership of pointing and demonstratives might be glimpsed. Several such patterns are described.

Chapter 3, 'Body-directed gestures', trains its focus in on a particular microcosm of pointing: gestures that speakers direct toward their own bodies. Body-directed gestures are a rich yet well-circumscribed ground for an exploratory investigation of form, function, and conceptual processes in pointing gestures. An analysis was carried out of body-directed gestures in a corpus of 40 one-on-one television interviews from the Tavis Smiley show. Based on this analysis, a typology is presented of three types of body-directed gestures— *self-points*, *body-points*, and *body-anchors*. Each type is considered in turn, with a focus on basic questions about how, when, and why body-directed pointing gestures are produced. In addition to addressing a basic descriptive blind-spot in the gesture studies literature, then, the study bears on theoretical questions of much more general concern. Discussion focuses on several issues that crosscut the three observed types of body-directed gestures, including variety in their form and function, the conceptual processes involved in their production, and their relationship to body-directed

analogues in sign languages.

Chapter 4, 'Nose-pointing in Yupno', describes a previously undocumented deictic gesture in use by the Yupno, an indigenous group of Papua New Guinea's Finisterre range. The gesture consists of a distinctive facial action— a scrunching together of the nose and lower brow— that is selectively layered on top of head-pointing. More than 80 examples of the gesture were identified from a corpus of high-definition field recordings made in the summer of 2009, and were analyzed for their form and contexts of use. Variation in the form of the gesture is described, as well as aspects of its coordination with other gestural articulators. Curiously, speakers also use the nose-pointing form in non-pointing contexts in the course of reference to small things or when foregrounding the property of smallness. This observation motivates the provisional proposal of a semantic kernel of *diminutiveness* underlying both deictic and non-deictic uses of the gestural form. Discussion focuses on the nature of the evidence in support of such a proposal, a consideration of the iconic roots of the gesture, and comparison of nose-pointing to lip-pointing.

Taken together the three chapters provide a composite characterization of the interplay of speech and pointing. Each chapter constitutes first, exploratory steps into different areas of theoretical interest. Many questions remain, and these are signposted in the chapters themselves, as well as in Chapter 5. At this point, it may be useful to back up and ask why it is that the co-patterning of speech and pointing might be of interest, in the first place, to researchers in the behavioral sciences. On the one hand, such patterns are of intrinsic interest. Human face-to-face interaction is, as many have argued, the central stage on which humanness emerges (e.g. Enfield & Levinson, 2006). Building a scientific

account of how interaction is organized, including both its species-wide characteristics and manifold cultural variation, is thus an important end in itself. Pointing is a small but no doubt critical part of such an account. Cognitive scientists may be drawn to the co-patterning of gesture and speech for yet other reasons. Patterns are not accidents; they contain clues about the processes that generate them. They contain clues, in this case, about how discourse is structured and about the kinds of conceptual processes involved in its production. They contain clues about the dynamics of attention and how it is managed multimodally in discourse. They contain clues about imagination and its unfolding in processes of real-time construal. And, finally, the patterns described in the present studies afford insights into a central puzzle of human meaning-making, the curious coupling of language and action.

References

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*, 415-419.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental psychobiology*, *40*(3), 293-310. doi: 10.1002/dev.10034.

Bühler, K. (1990 [orig. 1934]). *Theory of Language: The representational function of language*. Philadelphia: John Benjamins.

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 243-268). Mahwah, NJ: Lawrence Erlbaum.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, *17*(4), 463-489. doi: 10.1515/COG.2006.015.

Eco, U. (1976). *A theory of semiotics*. Bloomington, IA: Indiana University Press.

Enfield, N. J. (2001). "Lip-pointing": A discussion of form and function with reference to data from Laos. *Gesture*, *1*(2), 185-211. doi: 10.1075/gest.1.2.06enf.

Enfield, N. J. & Stivers, T. (Eds.). (2006). *Person reference in interaction: Linguistic, cultural, and social perspectives*. Cambridge: Cambridge University Press.

Enfield, N. J., Kita, S., & Ruiter, J. P. de. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722-1741. doi: 10.1016/j.pragma.2007.03.001.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.

Fauconnier, G. (1997). *Mappings in thought and language*. Cambridge: Cambridge University Press.

Frege, G. (1967 [orig. 1918]). The thought: A logical inquiry. In P. F. Strawson (Ed.), *Philosophical Logic* (pp. 17-38). Oxford: Oxford University Press.

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, *96*(3), 606-633. doi: 10.1525/aa.1994.96.3.02a00100.

Hanks, W. F. (2005). Explorations in the Deictic Field. *Current Anthropology*, *46*(2), 191-220. doi: 10.1086/427120.

Hare, B., & Tomasello, M. (2005). Human-like social skills in dogs? *Trends in cognitive sciences*, *9*(9), 439-44. doi: 10.1016/j.tics.2005.07.003.

Haviland, J. B. (1993). Anchoring, Iconicity, and Orientation in Guugu Yimithirr Pointing Gestures. *Journal of Linguistic Anthropology*, *3*(l), 3-45.

Haviland, J. B. (2000). Pointing, gesture spaces, and mental maps. In D. McNeill (Ed.), *Language and gesture* (pp. 13-46). Cambridge: Cambridge University Press.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495-514. doi: 10.3758/PBR.15.3.495.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S. (2003). Pointing: A foundational building block of human communication. In S. Kita (ed.), *Pointing: Where language, culture, and cognition meet* (pp. 1-8). Mahwah, NJ: Lawrence Erlbaum.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Levelt, W. J. M., Richardson, G., & Heij, L. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*, 133-164.

Levinson, S. C. (2006). On the human "interaction engine." In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39-69). Oxford: Berg.

Liszkowski, U., & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development*, *26*(1), 16-29. doi: 10.1016/j.cogdev.2010.10.001.

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, *32*(7), 855-878. doi: 10.1016/S0378-2166(99)00079-X.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: Chicago University Press.

McNeill, D. (2000). Introduction. In. D. McNeill (Ed.) Language and Gesture (pp. 1-10). Cambridge: Cambridge University Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: Chicago University Press.

Miller, G. A. (1982). Some Problems in the Theory of Demonstrative Reference. In R. J. Jarvella & W. Klein (Eds.), *Speech, Place, and Action: Studies in Deixis and Related Topics* (pp. 61-72). London: John Wiley & Sons.

Povinelli, D. J., Bering, J. M., & Giambrone, S. (2003). Chimpanzees' "Pointing": Another Error of the Argument by Analogy? In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 35-68). Mahwah, NJ: Laurence Erlbaum.

Schegloff, E. A. (1984). On Some Gestures' Relation to Talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 266-298). Cambridge: Cambridge University Press.

Sherzer, J. (1972). Verbal and nonverbal deixis: The pointed lip gesture among the San Blas Cuna. *Language in Society*, *2*, 117-131.

Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*. Philadelphia: John Benjamins.

Thao, T. D. (1984 [orig. 1973]). *Investigations into the origin of language and consciousness*. Hingham, MA: Kluwer.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Veà, J., & Sabater-Pi, J. (1998). Spontaneous Pointing Behaviour in the Wild Pygmy Chimpanzee (Pan paniscus). *Folia Primatologica*, *69*(5), 289-290.

Wagner, S. M., Nusbaum, H. C., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, *50*(50), 395-407. doi: 10.1016/j.jml.2004.01.002.

Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 171 - 215). Mahwah, NJ: Laurence Erlbaum.

Chapter 2. Pointing and the meaning of spoken demonstratives

2.1. Introduction

Demonstratives are special. They appear in every human language (Deissel, 1999), and are among the very first words children learn (Clark & Sengul, 1978). The English demonstratives— *this, these, that, those, here, there*— are extremely high frequency in adult conversation (Wu, 2004)[4]. Along with pointing gestures, demonstratives would appear to be among our most basic tools for coordinating attention in social interaction. Yet despite their primordial status, how demonstratives are actually used remains poorly understood.

Linguists have long analyzed demonstratives as encoding distance from the speaker (proximal = *this/ these, here*; distal = *that/ those, there*) (e.g. Lyons, 1977). The distance-based account continues to predominate in contemporary descriptive grammars of far-flung languages, as well as in foreign language textbooks. Unfortunately, the claims made by this distance framework are more often presumed self-evident than articulated with precision. On a naïve reading, the claim is that referents closer than a set metric distance (e.g. 6 feet) are referred to with proximal demonstratives, while referents beyond the set distance are referred to with distals. Such a naïve view might be termed the *metric* distance account (following Kemmerer, 1999). A more sophisticated reading,

[4] Just how frequent demonstratives are is difficult to estimate. Existing English-language corpora are largely based on written texts and, occasionally, on telephone conversations. Demonstratives are still frequently used in such settings, but their natural province is face-to-face interaction. Another issue complicating frequency estimates is that the form of the English distal demonstrative— *that*— also has a non-demonstrative use as a complementizer.

perhaps, is that referents which *in situ* are conceptualized as close to the speaker are referred to with proximals, while referents conceptualized as far from the speaker will be referred to with distals. This might be termed the *conceptual* distance view. A third possible reading is that demonstrative terms are spatial but only in an inherently relational sense: proximal terms are used when a referent is closer to the speaker than some other referent; distal terms are used when a referent is farther from the speaker than some other referent. This will be termed the *relational* distance account. These three distance-based accounts are by no means mutually exclusive, and, in fact, are readily combinable. One recent proposal, for instance, is that demonstratives are originally learned as corresponding to a basic metric distance distinction between *within* (proximal) and *beyond* (distal) peri-personal space, a perceptual distinction that only later give rise to more "removed uses" of demonstratives (Coventry et al., 2008, pg. 895).

In the last few decades, scholars in several fields have voiced doubt that demonstratives are best characterized as spatial in the first place. Several reasons for skepticism have been put forward. For starters, as Levinson (2004, pg. 109) has noted, there is no *a priori* reason that all demonstrative systems around the world should necessarily encode spatial distance contrasts, given that any number of other parameters are available in principle (e.g. gender). Others have argued from constructed examples that, on the one hand, proximal demonstratives are sometimes used felicitously for referents that are extremely *far* (e.g. "*this* planet," said of Mercury), while, at the same time, distal demonstratives are sometimes used felicitously for referents that are extremely *near* (e.g. "*that* tooth," said of a molar in the speaker's own mouth) (for discussion of such examples, see Talmy, 2000, pg. 25-6; Kemmerer, 1999, pg. 52;

Enfield, 2003, pg. 88; Piwek et al., 2008, pg. 699). A third reason for skepticism, perhaps, is that the distance-based account offers no straightforward explanation of the semantic relation between *exophoric* and *endophoric* uses of demonstratives. *Exophoric demonstratives* are those used to refer to entities in the immediate physical surround; *endophoric demonstratives* are those used to refer to entities in previous discourse or shared memory[5] (Halliday & Hasan, 1976). Yet many languages use the same words used for both endophoric and exophoric reference. In English, for example, the distal term (*that*) is preferred for *anaphoric* reference, or reference to previous discourse, while the proximal term (*this*) is preferred for *cataphoric* reference, or reference to upcoming discourse[6]. A distance-based account— at least of the *metric* variety— would seem to require that demonstrative words be polysemous or even homonymous, with disjoint endophoric and exophoric senses. A more parsimonious possibility is that a core semantic theme other than distance underlies both endophoric and exophoric uses of demonstratives.

For these and other reasons, researchers drawing on data from disparate languages have proposed that contextual parameters other than distance may better capture the semantic structure of demonstratives. Most notably, Hanks (1990, 2005, 2009) has developed an account of demonstratives in Yucatec Maya centered on the notion of access— perceptual, cognitive, and social— rather than distance. Simplifying somewhat,

---

[5] The tidiness of this distinction is perhaps questionable, but I will adopt it here for expository convenience.

[6] To my knowledge, no one has yet carried out a thorough-going study of cross-linguistic patterns in how exophoric contrasts (proximal/ distal) map onto endophoric contrasts (cataphoric/ anaphoric). But see Dixon (2003, pg. 83) for some discussion, as well as a sprinkling of data, suggesting the English mapping of proximal to cataphoric and distal to anaphoric may predominate.

his account argues that entities that are highly accessible are referred to with proximal forms; entities that are less accessible are referred to with distal forms. Variants of the idea that accessibility is the privileged parameter have now been explored in the analysis of demonstratives in Dutch (Piwek et al., 2008), Finnish (Ritva, 1996), Jahai (Burenhult, 2003), and Jordanian Arabic (Jarbou, 2010). It is beyond doubt that the accessibility account does a better job than the distance account of explaining a range of examples, including constructed examples like those presented above. One danger, however, is that the apparent explanatory power of the account could in fact stem from its over-generality. The account is plastic enough in principle to generate compelling *post hoc* accounts of why a given demonstrative form may have been used on a particular occasion of use. But its *predictive* power is less clear; it does not offer, for instance, a weighting of different subtypes of accessibility. In addition to the accessibility account, recent proposals have posited the centrality of other parameters, such as *here-space* (Enfield, 2003) and *control* (Brovold & Grush, *in press*)[7].

Many of the above studies have grounded their claims in observations of naturally occurring demonstrative usage. Naturalistic data have been critical in proving that demonstratives are— at least sometimes— used in ways that are at odds with a sheer distance account. What such data do not provide, however, is convincing evidence of general patterns in how different speakers in similar situations refer demonstratively. On reflection it is evident that there are some contexts in which, for example, use of one or

[7] Further parameters have been proposed in discussions of single terms in specific demonstrative systems. For instance, Küntay et al. (2006) have provided evidence of a Turkish demonstrative— *şu*— that encodes joint-attention status. Jungbluth (2003) has argued that the proper analysis of the Spanish three-term demonstrative system requires a dyad-oriented approach, rather than one based on sheer speaker-centered distance.

the other entity-referring demonstrative (*this* or *that*) is rigidly prescribed, and other contexts— perhaps more frequent— in which either form is in principle permissible (cf. Enfield, 2003, pg. 88, note 9). Isolated examples are thus minimally informative, and an important aim of future empirical investigations should be to describe *patterns* in demonstrative use.

Very few controlled quantitative studies of English demonstrative use have been reported in the literature. Bangerter (2004) and Bangerter & Chevally (2007) used a referential communication task to study the interplay of verbal and non-verbal strategies in coordinating attention. Participants identified human faces for each other from arrays set at different distances. In both studies it was found that the rates at which participants used pointing along with demonstratives changed with distance: the farther away the target, the less like the speaker was to use pointing-demonstrative combinations to refer to it. The studies did not analyze demonstratives produced without pointing, nor did they analyze patterns in which demonstrative— proximal or distal— was chosen. Wu (2004) had speakers of both English and Mandarin refer to puzzle pieces set at different distances and observed a trend of increasing use of distal forms for the farther pieces. A limitation of the study, however, is that co-speech gesture was not analyzed. Several other studies presenting quantitative data on demonstratives have likewise disregarded the co-involvement of gesture (e.g Clark & Krych, 2004), even as there is a growing recognition elsewhere that demonstrative use cannot be properly understood without attention to concurrent bodily action (Hindmarsh & Heath, 2000; Eriksson, 2009).

One recent experimental study has offered quantitative evidence that appears to vindicate a spatial account. Under the guise of a “memory game”, Coventry et al. (2008)

had speakers of both English and Spanish refer demonstratively to positions spaced at regular distances away from their body, ranging from 25 to 300 cm. Cards depicting colored objects (e.g. a red star) were placed at the different positions and participants were told that, according to the rules of the game, the only allowable references were of the form "*This* red star" or "*That* red star". Pointing was prescribed during all references. The data revealed in both languages a relatively abrupt transition point in demonstrative use, from proximal to distal forms, at approximately arm's length from the speaker— that is, at precisely the boundary between personal and extra-personal space. The authors interpret this finding as evidence that "demonstrative use corresponds with a basic distinction between near and far perceptual space" (pg. 889). A further finding of the study was that the point of transition between proximal and distal forms was shifted outward by giving participants a stick to point with, and the size of the shift was approximately equal to the length of the stick.

Results of this study seem to offer clear-cut evidence in two languages of the intrinsically spatial nature of demonstrative use. It remains possible, however, that demonstrative use is not based on space, but on some parameter that is very often confounded with space in the wild, as well as in Coventry et al.'s paradigm (c.f. Brovold & Grush, *in press*). One possibility along these lines is that demonstrative use is in part motivated by pointing affordances— that is, by whether or not referents afford pointing. The suggestion has *a priori* plausibility given the oft-mentioned co-incidence and functional link between demonstratives and pointing gestures (see Buhler, 1990 [orig. 1934]; Diessel, 2006, *inter alia*). As often as such a link is asserted, its particulars have not been worked out in any detail. How might differences in demonstrative usage be

linked to differences in pointing affordances? Wu (2004) has ventured in passing that one parameter potentially involved in exophoric demonstrative reference is "specificity" (pg. 42)[8]. Elaborating on this idea, we might propose that so-called "proximal" demonstratives are more likely to be used when a referent can be unambiguously pointed to by the speaker. "Distal" demonstratives, on this account, might be more likely to be used when pointing is more ambiguous. Here ambiguity is meant in a sheerly physical sense, leaving aside questions of mutual knowledge, discourse context, and other factors. If it is easy to discern which entity is being indicated given the nature of the spatial context, then the gesture is said to be unambiguous; if it is hard to discern which of several candidate entities is being indicated, then it said to be ambiguous. Note that, in the real world, as referents move farther away from the speaker, there is some tendency for them to become harder to point to. This is because manual, index-finger pointing projects a "cone" of possible referents toward a region of space. Note also that whether or not a potential referent is easy to point to— whether it is "demonstrable"— depends crucially on several spatial factors besides its sheer distance from the speaker, such as the size of the referent, its proximity to competitor referents, and, finally, on what resources the speaker has to point with. Coventry's et al.'s data are just as consistent with a demonstrability-based account as with a distance-based account, in both the case of the hand-pointing and tool-pointing conditions. In both cases, participants favored proximal

[8] Wu (2004) writes: "The asymmetry [between proximal and non-proximal forms] may also be seen as one of specificity or explicitness with which a referent is signalled" (pg. 42). It is not clear from this quote or the broader discussion whether Wu has a multimodal or strictly verbal sense of specificity in mind, however.

terms for objects that could be pointed to unambiguously and distal terms for objects that could not be pointed to unambiguously.

The present studies used a controlled referential communication task to explore patterns in demonstrative use. The studies had two main aims, one general and one specific. The first general aim was generate a corpus of spontaneous multimodal references involving demonstratives, produced by a large number of participants in a carefully controlled task setting. Such a corpus would provide much-needed quantitative data relevant to questions about how demonstratives partner with gesture, about how different demonstrative forms are preferred in different reference environments, and about levels of inter-speaker variability in demonstrative usage. The second more specific aim was to disentangle the effects of two contextual parameters— distance and demonstrability— on usage rates of proximal and distal forms.

2.2. Study #1

In Study #1 pairs of participants worked together, using both speech and gesture, to identify alien creatures from projected arrays. Distance and demonstrability were manipulated between participants to investigate their effects on demonstrative usage. Distance was manipulated by having the dyads sit at two different distances from the arrays, creating a NEAR condition and a FAR condition. Demonstrability was manipulated by having participants point by hand, in the low demonstrability condition, or with a laser pointer, in the high demonstrability condition. A laser-pointer was used because it effectively enables unambiguous reference at any distance. In contrast to the index finger, which projects a growing cone of possible referents, a laser pointer projects a narrow

vector. These manipulations thus yielded a 2x2 between-subject design with four conditions: NEAR-hand, FAR-hand, NEAR-laser, and FAR-laser.

Both a distance-based account and a demonstrability-based account would make the same predictions about demonstrative use in the two hand-pointing conditions, namely, that the proportion of proximal demonstratives would be higher in the NEAR condition than in the FAR condition[9]. Again, this is because when pointing by hand, all other things being equal, the ambiguity of pointing increases with distance. For the two laser-pointing conditions, however, the accounts make divergent predictions. The distance-based account would predict proportions identical to those found in the hand-pointing conditions. A demonstrability-based account could generate two slightly different predictions, but both distinguishable from the distance-based account. If demonstrability is the parameter that uniquely motivates demonstrative usage, then we should predict that, during laser-pointer conditions, the proportions would be globally higher than they are for the hand-pointing trials; crucially, a further prediction would be that there should be no difference between the two proportions in the NEAR- and FAR-laser conditions. High demonstrability, on this prediction, altogether neutralizes distance. An alternative prediction, consistent with a multi-parameter account in which distance and demonstrability both matter, would be that there should be an effect of demonstrability, as well as an additional effect of distance. On this prediction, dyads in the laser-pointing conditions would use a higher proportion of proximals overall than dyads in the hand-

[9] Though sharply distinguished here for expository purposes, the possibility that demonstrability affects demonstrative use does not exclude the possibility that distance also has effects. Both parameters might separately modulate demonstrative usage rates, perhaps with one being primary over the other.

pointing conditions, but the rates would be somewhat higher in the NEAR-laser condition than in the FAR-laser condition.

Both Study #1 and Study #2 adapted a referential communication paradigm originally used by Bangerter (2004) to investigate verbal and non-verbal strategies for coordinating attention in dialogue. One key difference, however, between the present study and Bangerter's original study is the use of alien creatures— or, Fribbles— in place of human faces. A distinct advantage of Fribbles is that they were designed for vision research to be of comparable novelty, complexity, and salience. Since salience has been previously implicated in demonstrative reference (Clark et al., 1983), it was important to equalize it across targets to the extent possible. Another important advantage of Fribbles over human faces is that they are less readily describable in familiar terms (e.g. "The blond guy with the beard") and thus should encourage the use of a demonstrative strategy.

2.2.1. Methods

2.2.1.1.Participants

86 UCSD students (43 female) participated voluntarily or in exchange for course credit. The majority of participants were monolingual native speakers of English. 18 participants were either bilingual from birth or had learned English later in childhood. All participants who had learned English later in childhood (8 participants) reported being exposed to the language by the age of 6 and, further, reported English as their dominant or co-dominant language at the time of the experiment.

Dyads were assigned to one of the four conditions based on a random order determined at the outset of the experiment. Excluded dyads (see below) were replaced at the end of the random order.

2.2.1.2. Materials

Stimuli consisted of novel creature-like objects (images courtesy of Michael J. Tarr, Carnegie Mellon University, http://tarrlab.cnbc.cmu.edu/). Four different types of novel objects were used: blue Fribbles, multi-color Fribbles, Greebles (pink), and Yadgits (multi-color). The objects— in all instructions referred to simply as "Fribbles" or "creatures"— were arrayed on Keynote slides in a cloud-like fashion, such that they did not form easily describable patterns**,** following Bangerter (2004). Each stimulus slide consisted of 12 Fribbles, all of the same type. Each dyad viewed 9 slides in all. They started with one practice slide of multi-color Fibbles, which was then followed by— in a different order for each dyad— two slides of blue Fribbles, two more slides of multi-color Fribbles, two slides of Greebles, and two slides of Yadgits. All stimulus slides consisted of only the 12 creatures on a plain white background (see Figure 2.1).

Participants sat in folding chairs in front of a projection screen. All instructions and stimuli were embedded in a Keynote slideshow and projected onto the screen with a Powerlite Home Cinema 6100 Projector. The dimensions of the projection were 56 inches wide by 42 inches high.

The task involved two roles, Finder and Recorder. Finders were given a stack of nine laminated 8.5 x 11 pages. Each page corresponded to a stimulus slide and included small pictures of the 12 Fribbles pseudo-randomly ordered in two columns along with

their names. Names were always a single digit number coupled with a single capital letter, e.g. “8A”. The Finders’ pages were placed on a music stand, which was positioned off to their left to allow unobstructed gesturing during the task. Small visual barriers were attached to the music stand so that Recorders could not see the Finders’ pages. Recorders were given corresponding pages, one for each slide, which were three-hole punched, and placed in a binder. The Recorders’ answer pages had only the pictures of the Fribbles and blank lines on which to write in the creatures’ names during the task. As a precaution, the positions of the small Fribble pictures were shuffled so as not to appear in the same places on the Finder's and Recorder's pages. To the upper right-hand corner of each one of the Recorder's pages was attached a numbered list of eight Fribble names, which specified the order in which the Fribbles were to be identified for each slide.

2.2.1.3. Procedure

After giving consent to participate and to be videotaped, participants entered the experiment room and were invited to take a seat in either of the two chairs. This determined their roles in the experiment, with Finders always on the left and Recorders always on the right. Exceptions to this self-assignment procedure were as follows. All left-handed participants were guided to the Recorder's seat, so that Finders would be uniformly right-handed. Handedness was covertly monitored by the experimenter at the time consent forms were signed. Further, participants who were known— or guessed— to be not native speakers of English were guided to the Recorder's seat in a similar way. Experimenter intuitions about whether or not a participant was a native speaker proved correct in every case, as revealed by a post-study language questionnaire.

Instructions were given on a Keynote presentation displayed on the projection screen. It was explained that participants would be working together at the Fribble family reunion to separate "real" Fribbles from "lookalikes" who were attempting to crash the party. Each slide consisted of eight real Fribbles that needed to be identified, as well four lookalikes. A name-list attached to each of the Recorder's binder pages listed the names of the eight real Fribbles in the order in which they were to be identified. At the start of a trial (slide), the Recorder was instructed to read aloud the name of the first Fribble on the name list. The Finder would then identify the Fribble for the recorder on the projection screen. After the Recorder had successfully recorded the name of the first real Fribble on the answer page, he or she would announce the name of the second real Fribble on the list. Across participants, the eight target Fribbles were the same for each slide, but were identified in a different, randomly generated order for each of the 32 dyads.

Participants were asked to stay seated, and were invited to interact in any way they would like to accomplish the task. Though gesture was not explicitly mentioned, the instruction slides included a photograph of model participants pointing. For the two laser-pointing conditions, the model Finder in the photograph was shown pointing the laser at the slide; for the two hand-pointing conditions, the model Finder was shown pointing to the slide with her right hand, index finger extended. The text accompanying these photographs did not draw attention in any way to the strategy depicted, only to the fact that the task involves distinct roles. We included this photograph for two reasons. First, in pilot studies, many participants in the hand-pointing conditions did not point, on the assumption— which they made explicit in debriefing— that pointing was not allowed. Second, the mere availability of the laser-pointer constitutes a strong suggestion about

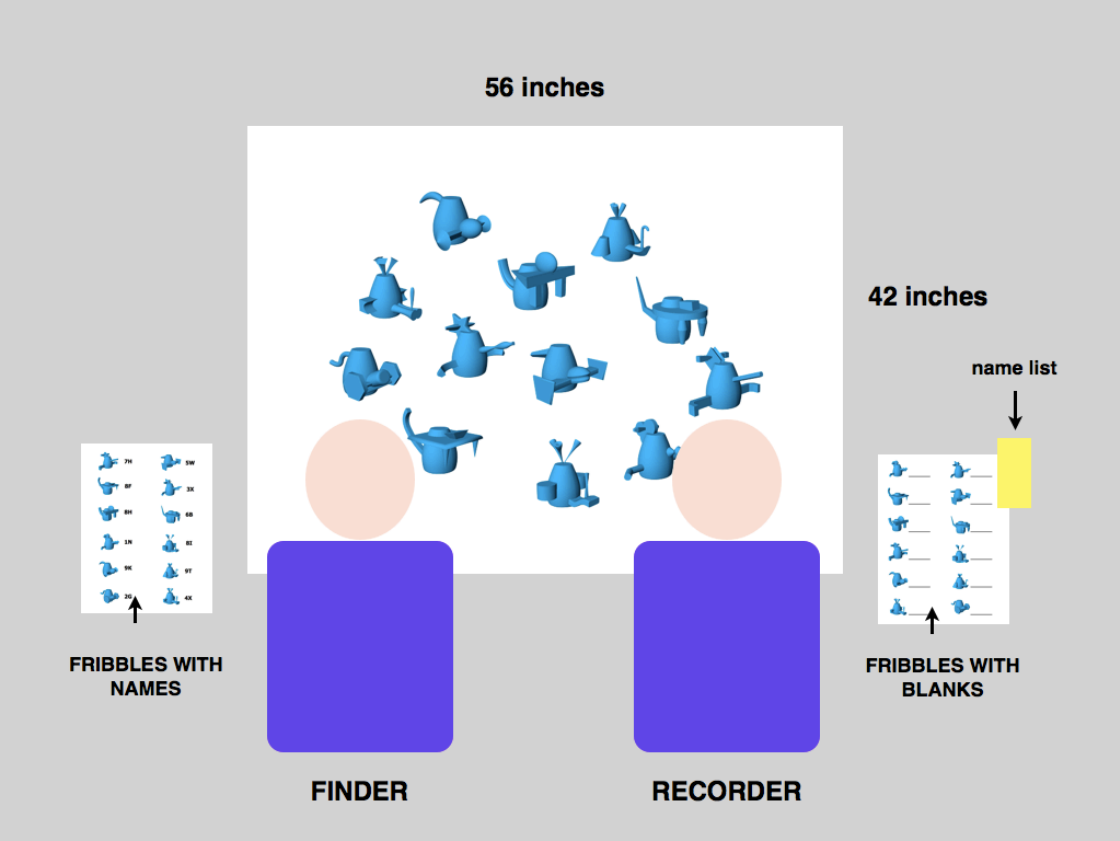

Figure 2.1: The set-up of the referential communication task used in both studies, which was adapted from Bangerter (2004).

how the task should be done, and in order to make the conditions more comparable we wanted to include a covert suggestion that pointing was a useful strategy in all conditions. That the suggestion was subtle is evidenced by the fact that more than once participants still asked during the practice round whether or not pointing was permitted. Instructions in the four conditions were identical except for the manipulation of this photograph. After the instructions were given, a practice slide was carried out, after which the experimenter left the room and the dyad proceeded through the eight stimulus slides, identifying a total of 64 target Fribbles in all (not including the practice targets).

In the NEAR conditions, participants' chairs were placed at a distance of 47 inches

from the projection screen, out of arms reach; in the FAR conditions, participants chairs were placed at a distance of 77 inches from the screen. For laser-pointing dyads, the pointer was placed on the Finder's stand before the start of the experiment.

2.2.1.4. Data collection and analysis

Sessions were video-recorded with a Canon HV20 HD digital camera. Tapes were digitized in Final Cut Pro, rendered into clips of separate slides (8 clips per dyad), and analyzed using ELAN video annotation software (available online: http://www.lat-mpi.eu/tools/elan/).

Data from 22 participants (11 dyads) were excluded from analyses for different reasons. One dyad was eliminated because the Finder stood up and approached the screen on more than one trial, in clear violation of the instructions. Another dyad was eliminated because the Finder guessed in debriefing that we might be interested in the word *that*. Two dyads were eliminated because the participants did not give us permission to view the videotapes. And, finally, seven dyads were eliminated from the laser pointer conditions because they talked extremely infrequently, producing six or fewer demonstrative (+pointing) references over the course of the experiment. Ultimately, data from 64 participants (33 male; mean age= 20.7 years old), forming 32 dyads, eight in each of the four conditions, were included in the final analyses.

Each use of a demonstrative— *this, that, here, there, these, those*— by either participant was annotated, along with certain features of the context of use. Most importantly, it was noted whether or not the speaker was pointing concurrently. Also noted was whether the demonstrative was used in reference to a Fribble on the projection,

a Fribble not on the projection screen, or to something besides a Fribble (e.g. the pen the Recorder was using). Only demonstratives used in reference to Fribbles were analyzed extensively, and all quantitative data below include exclusively references to Fribbles. Uses of the plural demonstratives *these* and *those* were rare, and were not analyzed further. Uses of *that* as a complementizer (e.g. "The one **that's** on the right") were not annotated; uses of the *there* of existence (e.g. "**There's** a blue cone on its head") were annotated but not analyzed further. The manner demonstratives *like this* and *like that* were annotated but were considered as distinct demonstrative constructions and not lumped in with others. Pointing in the hand-pointing trials was operationalized as arm and finger extension towards the projection screen. Pointing in the laser-pointing trials was operationalized as visible red light from the pointer directed at the screen. For the hand-pointing dyads, reliability was assessed by having a second coder judge the presence or absence of pointing during two trials for each dyad (i.e., 25% of the total trials). Agreement was almost perfect (94.3%; Cohen's Kappa= 0.887). The judgments of the primary coder were retained for the purposes of quantitative analysis. All proportions obtained were arcsin-transformed prior to statistical analysis.

2.2.2. Results and discussion

2.2.2.1. Qualitative description of data

The task was carried out in the same basic way across the four conditions. Dyads identified the target Fribbles one at a time according to the following steps. First, the Recorder would announce the name of the Fribble (e.g. "The next one is 5W"). Finders would very often confirm the name by repeating it aloud. They would proceed to locate

the correct Fribble on their own answer page on the music stand, then find it visually on the projection screen, and, finally, begin to identify the Fribble for the Recorder. The launch of the identification sequence would often include *it* (e.g. "**It** has the star nose"), *this* (e.g. "**This** one has the star nose"), *that* (e.g. "**That** one has the star nose"), or a repetition of the Fribble's name (e.g. "**5W** is the one with the star nose"). Which demonstratives were used at the start of identification sequences differed from Finder to Finder, but not in a systematic way across conditions. Such demonstratives were considered *endophoric*: they referred to the Fribble as part of a sequence of Fribbles that had to be identified. As the identification process proceeded, the Finder would then coordinate the sequence referent—i.e. the current targeted Fribble— with one of the 12 visible creatures on the projection screen. Endophoric (sequence-referring) demonstratives and exophoric (projection-referring) demonstratives were analyzed separately (see below). Following the successful identification of the target on the screen, Recorders then needed to find the corresponding Fribble on their own answer pages in order to write in the name in the blank. Recorders would often conclude by re-coordinating the projection referent with the sequence referent, either aloud or quietly to themselves (e.g. "Okay, so that one's 5W").

Strategies for identifying the projection referent differed systematically in the different conditions. In the NEAR-hand condition Finders used demonstratives, feature descriptions, deictic gestures, and iconic gestures to identify Fribbles on the projection, and they often used these resources in combination. This mix of strategies is evident in the following example (see also Figure 2.2):

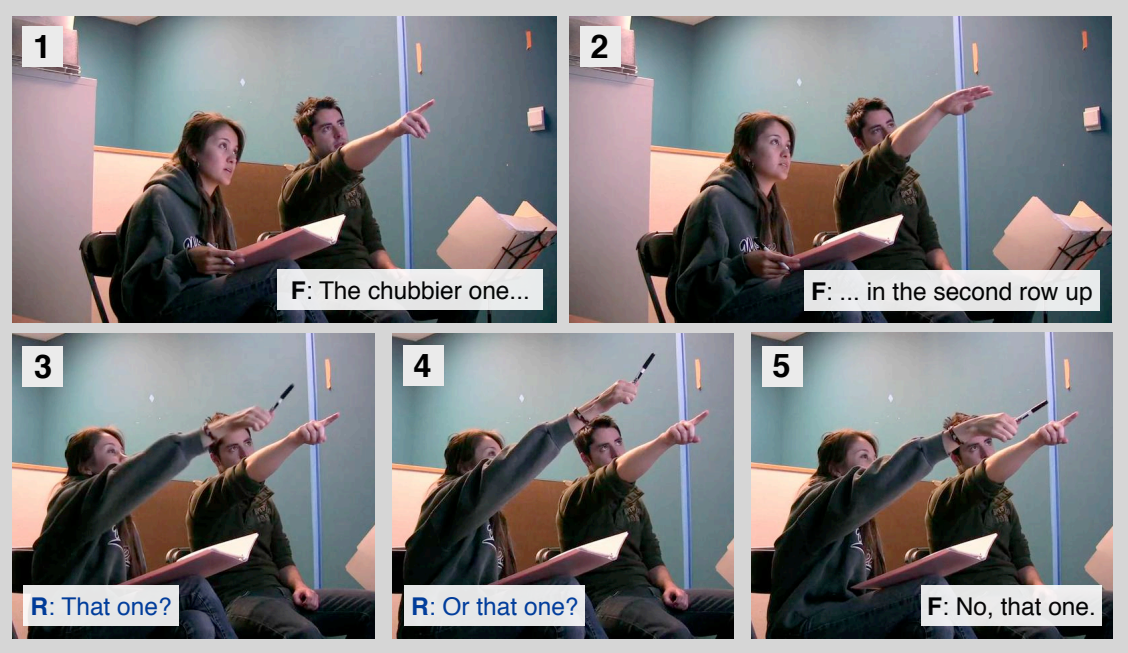

Figure 2.2: A sequence from the NEAR-hand condition that illustrates the use of different strategies for identifying targets on the projection screen. The transcript is given as example (1) in the text.

Example (1): NEAR-hand condition

Recorder: Um, 3Z.
Finder: 3Z. Alright, it's the one right next to the chubbier one with the uh-uh with the goofier mustache. It's like in the second row up.
Recorder: That one? *[pen pointing]*
Finder: Yeah.
Recorder: Or that one? *[pen pointing]*
Finder: No, that one. *[hand pointing]*
Recorder: That one. *[pen pointing]*

Finders did not use demonstratives on every trial, and there was considerable inter-participant variation in how often Finders used a demonstrative as part of their referring expressions.

In the FAR-hand condition, the rate of exophoric demonstrative usage dropped off considerably — indeed, several finders never used demonstratives (+ pointing). Most

relied instead on a combination of feature descriptions, location descriptions, and iconic gestures, as evident in the following example:

Example (2): FAR-hand condition

Recorder: 3Q.
Finder: 3Q. The gun one we just did… up to the left diagonally. It's got the square coming off the back with the kickstand and the two cylinders on the side going forward holding like a ball.
Recorder: The square one. Okay. So- so- from the one we did…
Finder: Mm-hm.
Recorder: The- the to the left- it's to the top left.
Finder: Yeah, yeah.
Recorder: Okay.
Finder: It kinda looks like a dragon, picking up a dragon or something.
Recorder: That's 3Q.

In both of the hand-pointing conditions, participants occasionally attempted the strategy of carving the array into numbered rows and columns. This strategy met with mixed success, however, because of designed irregularities in the arrays. In every case participants also found it useful to supplement with other strategies.

In the laser conditions, Finders used almost exclusively laser pointing accompanied by demonstrative reference. Some Finders, however, did not regularly speak at all. Instead, the Recorder would say the name, the Finder would point at it wordlessly, and the Recorder would move to the next one. Finders who produced fewer than 6 demonstrative (+pointing) references over the course of the 64 targets were excluded from the final data analysis, as detailed above (see Section 2.1.4). Finders who did speak in the laser conditions were in most cases very efficient, as in the following example:

Example (3): FAR-laser condition

Recorder: Um, 2D.
Finder: And that's this one. *[laser pointing]*
Recorder: Cool.

Finders also occasionally provided minimal feature descriptions in addition to pointing out the target.

Recorders participated to different degrees in the identification of the Fribbles. Some Finders adopted a strategy in which they first described the Fribble in detail based on the small image on their answer sheet before even looking up, thereby effectively recruiting the Recorder to help locate the target on the projection screen. In other cases Finders would only appeal to the Recorder for help when having difficulty, as in the following example:

Example (4): FAR-laser condition

Recorder: 3M.
Finder: 3M. Uh, kind of could be this one. *[laser pointing]*
But the things look kind of fatter.
Recorder: Oh- oh is it straight?
Finder: Yeah.
Recorder: It's- all the way down. Is it that one? *[hand pointing]*
Or no?
Finder: This one? *[laser pointing]*
Recorder: Yeah.
Finder: No. Cause the things are on the bottom.
Recorder: Oh, okay.
Finder: Is it-?
Recorder: That one. *[hand pointing]*
Finder: This one yeah. *[laser pointing]*

As is evident in this example, as well as in example (1), Recorders also used demonstratives and pointing gestures— sometimes with their pens— both to suggest and

confirm targets. Different Recorders participated at starkly different rates, making any quantitative analysis of Recorders' demonstratives difficult.

2.2.2.2. Quantitative results

The task elicited a large number of demonstrative references (n= 2463). As expected given the design of the task, the majority of these demonstratives (1888, or 76.7%) were produced by the Finder.

In their demonstrative references to Fribbles, Finders preferred *this* over *that* when pointing (637/904, or 70.1%) and *that* over *this* when not pointing (198/338, or 58.7%). For Finders in Study #1, presence of pointing and choice of demonstrative form were not independent ($\chi^2 = 88.61$, df=1, p= <0.0001), a pattern which is also evident in each of the four conditions when considered separately.

Because our particular interest was in exophoric demonstratives, we limited further quantitative analyses to cases in which the Finder was referring demonstratively while also pointing (n= 1307). Finders only very rarely referred demonstratively to a Fribble on the projection screen without simultaneously pointing to it. They did, however, often refer to Fribbles endophorically without pointing. Such endophoric references were used when Fribbles were construed as part of a task sequence (e.g. "**This** next Fribble" or "**That** last Fribble") as described above.

Demonstrative references (-pointing) to Fribbles did not systematically differ by condition, but demonstrative references (+pointing) did. Finders used a higher rate of demonstratives (+pointing) per slide in the NEAR-hand condition than in the FAR-hand condition (NEAR-hand: mean= 3.55 [std. error= 0.74]; FAR-hand: mean= 1.38 [std. error=

0.39]), and more often still, at roughly equal rates, in the two laser conditions (NEAR-laser: mean= 7.48 [std. error= 1.21]; FAR-laser: mean= 8.02 [std. error= 0.81]). We analyzed these mean rates using a two-way independent samples ANOVA with distance (NEAR, FAR) and pointing type (hand, laser) as between-subjects factors. The analysis revealed a main effect of pointing type ($F(1,31)= 39.64$, $p< 0.0001$), but no main effect of distance ($F(1,31)= 0.95$, $p= 0.338$) and a trending but non-significant interaction effect ($F(1,31)= 2.59$, $p= 0.119$). This observed main effect replicates one of the findings of Bangerter (2004) that speakers use demonstratives (+pointing) less as they become less able to point unambiguosly.

The above analysis clearly demonstrates a difference by condition in the extent to which demonstratives (+pointing) were used as a reference strategy. Did the conditions also differ in terms of which *forms* of demonstratives were used? To address this question, we compared the mean proportions of proximal demonstratives used in each of the four conditions. (Note, however, that three dyads from the FAR-hand condition were excluded from this analysis because they produced fewer than 6 demonstratives, making their proportions less statistically meaningful.) Finders used a higher proportion of proximal demonstratives (*this* and *here*) in the NEAR-hand condition (mean= 0.56 [std. error= .09]; median= .58) than in the FAR-hand condition (mean=0.37 [std. error= .22]; median= .17). In the laser-pointing conditions, Finders used proximals in higher proportions than in the hand conditions, and only marginally more in the NEAR (mean= 0.75 [std. error= .07]; median= .76) than in the FAR condition (mean=0.68 [std. error= .11]; median= .81) (see Figure 2.3). We analyzed these mean proportions using a two-way independent samples ANOVA with distance (NEAR, FAR) and pointing type (hand,

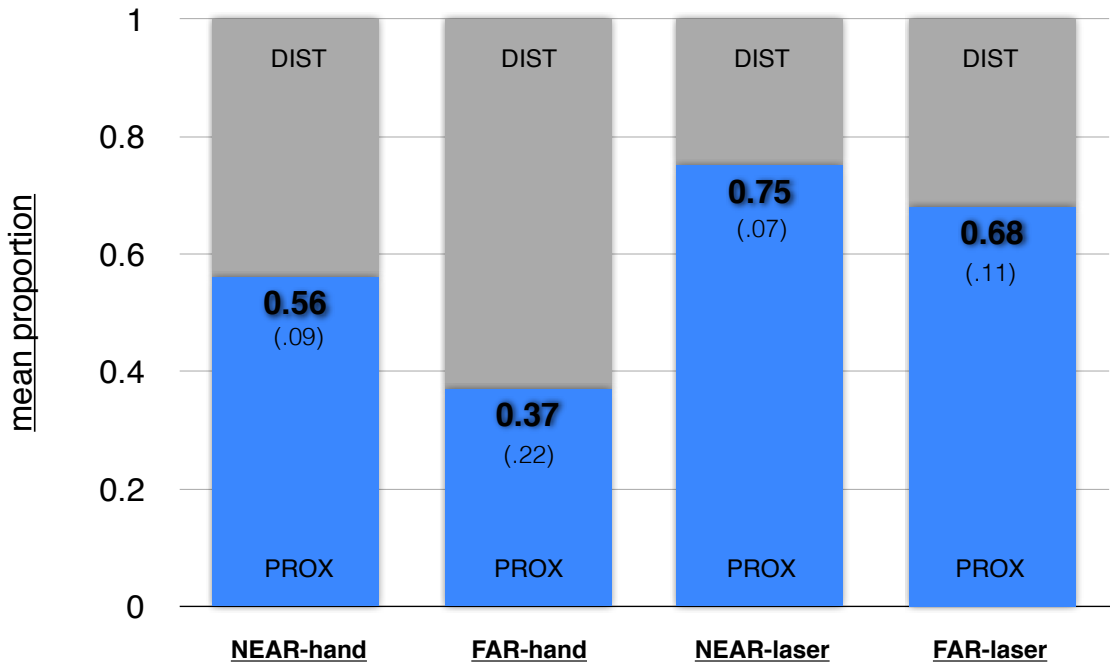

Figure 2.3: Mean proportions of proximal demonstratives (+pointing)— *this* and *here*— used by Finders in the different conditions of Study #1.

laser) as between-subjects factors. The analysis revealed a significant main effect of pointing type ($F(1,28)= 4.95$, $p= 0.035$), but no main effect of distance ($F(1,28)= 1.03$, $p= 0.32$) and no interaction effect ($F(1,28)= 1.03$, $p= 0.32$)[10].

Evident above is a high level of inter-participant variability, both in the rates of demonstratives used and in the patterns of forms used. In the NEAR-hand condition, Finders used between 1.38 and 7.88 demonstratives (+pointing) per slide; in the FAR-hand condition, Finders used between 0.00 and 3.38 per slide. In the laser-pointer conditions, Finders also varied in their rates of demonstrative (+pointing) usage (NEAR-laser: 1.75 to

[10] Conducting the ANOVA analysis without first arcsin-transforming the proportions yields the same pattern of results, but with a slightly higher p-value of 0.052. The same analysis comparing only proportions of entity-referring demosntratives *this* and *that*— excluding the place-referring forms *here* and *there*— yields the same pattern of results, but with a p-value of 0.02.

13.38; FAR-laser: 3.25 to 10.50). Proportions of proximal usage across different speakers were also highly variable (NEAR-hand: 0.07 to 0.89, inter-quartile range= 0.18; FAR-hand: 0.00 to .90, inter-quartile range= 0.79; NEAR-laser: 0.43 to 1.00, inter-quartile range= 0.26; FAR-laser: 0.04 to 1.00, inter-quartile range= 0.40). Such variability complicates statistical analysis, but also perhaps provides a crucial datum about the nature of demonstrative use. This point will be elaborated in discussion below.

The above results provide evidence of coarse-grained differences between the four conditions in the rates at which demonstratives were used, as well as in the proportions of proximal terms used. A further question, however, is whether demonstrative use was also modulated by more fine-grained distance contrasts. Since Finders are seated on the left side of the projection screen (see Figure 2.1), they were not equally distant from each target. We can thus also investigate whether they keyed on distance target-by-target, referring to closer targets differently from farther targets. We addressed this question *post hoc* by determining the distance along the projection plane (in inches) of each of the 64 target Fribbles from an idealized reference point, which was calculated to be an estimate of the average nose position of the Finder. Target distances ranged from 2.2 inches for the nearest Fribble to 32.0 inches for the farthest Fribble. For each of the targets, we determined whether or not the Finder used a proximal form to refer to it, a distal form to refer to it, or both[11]. We then collapsed this information across Finders for each of the four conditions, yielding for each target the total number of

[11] This analysis is slightly noisy for at least one reason. The video recording does not allow us to see *where* on the screen speakers are pointing, only *whether* they are pointing. It is likely that Finders occasionally made mistakes, pointing to the wrong creature on the projection. However, overall, participants did not consider the task particularly challenging and most dyads carried it out with minimal back-tracking to previous targets.

Finders who referred to it with a proximal and the total number of Finders who referred to it using a distal demonstrative.

For a first analysis, we calculated the correlation between each Fribble's distance from the idealized reference point and the number of times it was referred to demonstratively. In the NEAR-hand condition, this correlation was highly significant ($r= -0.54$, $r^2= 0.29$; $t(62)= -5$, two-tailed $p= < 0.0001$), as it was in the FAR-hand condition ($r= -0.43$, $r^2= 0.19$; $t(62)= -3.8$, two-tailed $p= 0.0003$). In the NEAR-laser condition, by comparison, there was no such correlation ($r= 0.10$, $r^2= 0.01$; $t(62)= 0.79$, two-tailed $p= 0.433$); nor was there a correlation in the FAR-laser condition ($r= 0.259$, $r^2= 0.0007$; $t(62)= 0.2$, two-tailed $p= 0.842$). Participants in the hand-pointing conditions were thus less likely to use demonstratives for targets farther away on the projection plane; participants in the laser-pointing conditions were equally likely to refer to Fribbles with demonstratives regardless of their location on the screen. This result provides a different, more fine-grained replication of Bangerter's (2004) finding that the use of demonstratives (+pointing) hinges critically on the ability to point out referents.

The analysis of mean proportions of proximals by condition showed a main effect of pointing type. A further question, however, concerns the nature of the effect. Did use of the laser merely drive up the baseline rate of proximal usage, or did it neutralize distance entirely? To explore this question we collapsed data from both laser conditions to determine whether speakers used proximal demonstratives in different proportions when referring to near versus far Fribbles. For this analysis we eliminated the middle 50% of targets, considering only the nearest 16 Fribbles and the farthest 16 Fribbles. This constitutes a relatively powerful within-subjects test of whether or not Finders keyed on

distance at all when using the laser pointer. The mean proportion of Finders' proximal demonstratives was 0.74 (std. error= 0.09; median= 1.00) for the nearest 16 targets and 0.66 (std. error= 0.075; median= 0.782) for the farthest 16 targets. A paired-samples t-test revealed that this difference is indeed significant ($t(15)= 2.76$, two-tailed $p= .015$). 10 out of the 16 Finders exhibited the pattern, 3 had identical proportions for the two sets of targets, and 3 Finders showed the opposite pattern. This finding suggests that, despite the use of the laser pointer, distance modulated proximal usage rates to a small but nonetheless measurable degree. Unfortunately, corresponding tests for the hand-pointing conditions were not statistically interpretable because of an absence of data, as participants rarely used demonstratives to refer to the farthest Fribble targets when pointing by hand.

2.3. Study #2

The results of Study #1 suggested that demonstrability modulates speakers' choice of demonstrative forms. Finders used proximal forms in higher proportions when they could point unambiguously (laser conditions) than when they could not (hand conditions). Distance was also found to have an effect— albeit small— independent of the effect of demonstrability. Even though they could point unambiguously to all targets, Finders in the laser-pointer conditions still used a smaller proportion of proximals for the farthest Fribbles. In the second study, we sought to replicate the main findings of Study #1 while also investigating whether the effect of demonstrability was robust to differences in the task.

One specific question raised by Study #1 was the extent to which task role—Finder or Recorder— may have modulated the use of demonstratives independently of the two isolated contextual factors, distance and demonstrability. The question could not be investigated thoroughly in the first study because only Finders produced demonstratives in sufficient quantities to allow for meaningful statistical comparisons. Task role could have modulated demonstrative rates in several ways. First, it is possible that Finders used high rates of proximals in part because of a knowledge asymmetry built into the task: Finders know the correct locations of the Fribbles on the slide, while Recorders do not. Second, Finders' references to the Fribbles are most often initial references, while Recorders' references are most often subsequent references. We thus wondered whether Recorders would have exhibited a different pattern of demonstrative usage from Finders had they been given occasion to use more demonstratives.

To explore these issues, in Study #2 we ran an additional 10 dyads in the NEAR-hand condition with a small but important change: we also provided the Recorder with a laser pointer. We reasoned that this manipulation would induce participants in both roles to use demonstratives (+pointing) in large quantities, instead of only the Finder, as was the case in Study #1. If distance and demonstrability alone shape demonstrative use, then the observed patterns for the Recorder in Study #2 should be comparable to those of the Finder in the NEAR-laser condition of Study #1. At the same time, the observed patterns for the Finder in Study #2 should mirror those of the Finder in the NEAR-hand condition of Study #1. However, if other factors are important, a number of different patterns could emerge.

2.3.1. Methods

2.3.1.1. Participants

24 UCSD students (14 women) participated in the study in exchange for course credit. 16 of the participants were monolingual speakers of English; 8 were bilingual to various degrees but with English listed as their dominant language.

2.3.1.2. Materials

The materials were identical to those used in Study #1.

2.3.1.3. Procedure

The instructions and procedure were identical to those for Study #1 with one important exception: immediately before the practice round the experimenter handed a laser pointer to the Recorder, explaining that it was available for use but that it could not be transferred to the Finder.

2.3.1.4. Data collection and analysis

Video was captured and prepared in the same way it was for Study #1. Data from one dyad was eliminated because one of the participants did not give us permission to view the videotape; data from another dyad was eliminated because in the post-study language questionnaire the Finder identified a language other than English as his dominant language. Ultimately, data from 20 participants (13 women; mean age= 20.78 years old), resulting in 10 dyads, were included in the final analyses.

Analysis of the video proceeded in the same way it did for Study #1. However, two additional guidelines were introduced into the video annotation procedure to address strategies and patterns particular to Study #2 (for fuller description, see section 3.2.1). First, presence of laser-pointing was no longer operationalized simply as visible red light from the pointer. Several Recorders in Study #2 kept the laser pointer turned on even while oriented, not to the screen, but to their own answer sheets. To deal with this complication, we changed the operationalization to require both visible red light *and* visual orientation to the projection screen. Another unforeseen issue was that the view of the Recorder's laser was occasionally obscured by the binder they were holding. In such cases the video analyst inferred the presence or absence of pointing from the context of the interaction, usually with little difficulty. Lastly, we further chose to annotate whether the Recorder was pointing to the screen during the *Finder's* demonstrative references. Reasons for this additional annotation will be discussed below.

2.3.2. Results and discussion

2.3.2.1. Qualitative description of data

Participants in Study #2 carried out the task in a similar way to participants in the first study. The Recorder started by announcing the name of the first Fribble. The Finder confirmed the name, and then coordinated a sequence reference to the to-be-identified target Fribble with a reference directed at the projection. Projection references involved the same combination of strategies used by Finders in the NEAR-hand condition of Study #1, including pointing, demonstrative reference, location descriptions, and iconic gesture. An extremely common pattern was that the Finder would identify the target on the screen

through a combination of strategies, and the Recorder would confirm the identification by pointing to it with the laser. This confirmation strategy is evident in the following example:

Example (5): Study #2

Recorder: Uh, 8F.
Finder: 8F. Um, this bottom leftish one**.** [*hand-pointing*]
Recorder: This one? [*laser-pointing*]
Finder: Yeah.

Additional strategies also emerged. Strategy choice was negotiated dyad by dyad, and was by no means the autonomous choice of the Finder. In some dyads, the Finder directed the Recorder verbally to the target and then confirmed when they had successfully arrived, as in the following example:

Example (6): Study #2

Recorder: 8J.
Finder: Okay, um start at the top. Go down one to the left.
Recorder: [*laser-pointing*]
Finder: There it is.
Recorder: Okay.
Finder: That's 8J.

Overall the set-up of Study #2 invited markedly more participation from the Recorder.

2.3.2.2. Quantitative results

The task elicited a large number of demonstratives (n= 1311), this time by both the Finder (n= 726) and the Recorder (n= 585).

Finders and Recorders used *this* and *that* in approximately equal proportions

when pointing (Finder: *this*= 99/222, or 44.6%; Recorder: *this*= 147/301, or 48.8%) but strongly favored *that* when not pointing (Finder: *that*= 239/282, or 84.8%; Recorder: *that*= 106/127, or 83.5%). Chi-square tests reveal that, as in Study #1, pointing and choice of demonstrative where not independent, both for the Finder ($\chi 2 = 52.86$, df=1, p= <0.0001) and the Recorder ($\chi 2 = 39.08$, df=1, p= <0.0001).

The mean rates of demonstratives (+pointing) were similar for both the Finder (3.86 per trial, std. error= 1.05) and Recorder (4.35 per trial, std. error= 0.57). The mean proportion of proximal (*this, here*) demonstrative (+pointing) uses by Finders was 0.37 (n= 8, std. error= 0.12; median= .39) and by Recorders was 0.53 (n= 10, std. error= 0.12; median= 0.63)[12]. (Note that two Finders were excluded from the analysis because they produced fewer than 6 demonstratives over the course of all trials.) Though the general pattern echoes the findings of Study #1 that high demonstrability leads to higher proportions of proximal demonstratives, a two independent samples t-test of the difference between the mean proportions yields a non-significant result (t(16)= -0.85, two-tailed p= 0.407). Further, while the Finders' mean proportion of proximal forms appears lower than that found in NEAR-hand condition of Study #1, the difference is not significant (t(14)= -1.41, two-tailed p= 0.18), nor is the difference between Recorders' mean proportion of proximal forms and that of the Finder in the NEAR-laser condition of

[12] It may seem simplistic to lump entity-referring (*this*) and place-referring demonstratives (*here*) together in these analyses, given that they clearly have different semantics on the whole. In fact, though, *here* has two distinguishable uses in English. The first is in the sense of a place that encompasses the speaker (e.g. *Here in San Diego,* said by a speaker in San Diego); the second is in the sense of a place that the speaker is concurrently pointing to (e.g. *Here in Tokyo*, said pointing to Tokyo on a map) (see Miller, 1982, pg. 71 for discussion of the difference.) Only the second sense of *here* is used in this task.

Study #1 ($t(16)= 1.59$, two-tailed $p= 0.13$). The lack of significant results are difficult to interpret, however, as they may result from a lack of statistical power.

As in Study #1, inter-participant reliability was high, both in the rates at which demonstratives were used and in the proportions of proximal forms used. Finders' proportions of proximal usage ranged from 0.00 to .763 (inter-quartile range= 0.62); Recorders' proportions of proximal usage ranged from 0.00 to .967 (inter-quartile range= 0.57). Interestingly, by some informal measures, variability was higher in Study #2 than in Study #1. The inter-quartile range for the Finder (0.62) was larger than that for the Finder in the NEAR-hand condition of Study #1 (0.18); the inter-quartile range for the Recorder (0.57) was larger than that for the Finder in the NEAR-laser condition of Study #1 (0.26). Strikingly, two Recorders in Study #2 never once used a proximal form (0/34 and 0/42). Of all 16 Finders in Study #1 who used the laser pointer, only a single speaker used proximal forms less than 40% of the time; in Study #2, 4 out of the 10 recorders did. Though these differences are hard to interpret, it is possible that the apparent increase in variability is related to variation in strategies used, in how the task was conceived by the participants, and in the kinds of circumstances in which demonstratives were used.

An unexpected but important finding of Study #2 was the emergence of a particular well-circumscribed context in which participants where nearly unanimous in their choice of demonstrative form. As illustrated in example 6 above, Finders in Study #2 very often verbally confirmed for the Recorder that they had alighted on the correct target, and very often did so using a demonstrative. Finders very frequently referred demonstratively to a Fribble that the Recorder was pointing to concurrently (n= 217), but did so using distal demonstratives in almost every case (*that*= 156*; there*= 59; *this*= 1;

here= 1). Such contexts occurred in every dyad (min= 3 occurrences; max= 39 occurrences). What's more, counted among these 217 instances are 53 cases in which the Finder him- or herself was *also* pointing to the target Fribble, yet they nonetheless chose distal demonstratives in every case (*that*= 34; *there*= 19; *this*= 0; *here*= 0). Note, importantly, that such references are clearly exophoric. Both participants are visually oriented towards a referent in the visible, physical surroundings. Nor is this merely an effect of first-mention versus subsequent mention. At times the Finder's demonstrative was produced immediately following a verbal reference by the Recorder (e.g. "This one?"), but in many cases— as in example 6— the Recorder points without speaking. A possible interpretation of this finding will be considered below.

Finally, we pooled data from the two NEAR conditions of Study #1 and from Study #2 to explore the robustness of the effect of demonstrability. We conducted a two-way independent samples ANOVA on the mean proportions of proximal terms used by the Finders in the NEAR-hand condition (Study #1), the Finders in the NEAR-laser condition (Study #1), the Finders (hand-pointing) in Study #2, and the Recorders (laser-pointing) in Study #2. Task set-up (Study #1, Study #2) and pointing type (hand, laser) were used as between-subjects factors. The analysis revealed no main effect of task set-up ($F(1, 33)= 2.22$, $p= 0.14$), a significant main effect of pointing type ($F(1,33)= 4.13$, $p= 0.05$), and no interaction effect ($F(1,33)= 0.44$, $p= 0.51$). This analysis provides evidence that the effect of demonstrability persisted in Study #2.

2.4. Discussion: Demonstrability, deictic force, and demonstrative meaning

A primary finding of the above studies is that, all other things being equal, the

"demonstrability" of targets— that is, how easy it is to point them out— affects whether speakers refer to them using proximal or distal demonstrative forms. Speakers used a greater proportion of proximal demonstrative terms— *this* and *here*— in the high demonstrability conditions, in which it was possible to refer unambiguously to targets by means of a laser pointer. In the wild, demonstrability and distance are very often entangled. Notwithstanding other factors, when pointing by hand, far targets are more difficult to point to than close targets. The present study thus offers a plausible motivation for the prevailing distance-based view, at the same time that it provides evidence for the importance of a non-spatial parameter.

An alternative explanation of the above result is that the laser pointer used in the present studies induced a change in body schema, with the effect that distant targets were experienced as closer. This explanation is plausible at first blush. It is well-known that body schema shifts occur in both humans and non-human primates with tool use (Maravita & Iriki, 2004). Importantly, however, the types of tools used in such studies— which include rakes, sticks, and golf clubs— differ from laser pointers in several key respects: first, they are rigid and thus provide the tool-user tactile feedback when their tips contact objects; second, they have a fully visible extension; and, finally, they are of a fixed length. Laser pointers, by contrast, do not provide tactile feedback, are invisible except for at the precise point at which they "contact" objects, and are of an indeterminate length. In fact, perhaps because of these features, laser pointers have been used extensively in clinical contexts as a tool to diagnose near and far neglect. For example, Berti & Frassinetti (2000) documented the case of a patient with near-space neglect. The patient performed poorly on line bisection tasks carried out in close peri-

personal space, both when bisecting by hand and when bisecting with a laser pointer. When executing the same line bisection tasks in far space with a laser pointer, the patient showed no impairment. Strikingly, when subsequently performing the same task in far space with a rigid tool (in this case a 100cm stick), the near-space impairment resurfaced. The authors explain these results on the idea that use of the rigid tool induced an expansion of the body schema and thus an expansion of the near-space neglect. Note, critically, that the laser pointer induced no such expansion. In short, no current evidence exists to support the idea that the body schema might be plastically expanded through use of a laser pointer, and the best available evidence suggests it does not. Further research will be required, of course, to confirm that the construct of demonstrability is what best explains the differences observed above between the hand-pointing and laser-pointing conditions. Other interpretations may indeed be possible.

Demonstrability in the present studies was operationalized as the means by which a speaker can point to objects, either with the hand (low demonstrability) or with a laser-pointer (high demonstrability). Other factors not investigated here also affect how well targets afford pointing. Targets are easier to point to, for example, when they are spaced apart from competitor targets, or when they are larger. Future studies will be required to manipulate these other determinants of demonstrability as well, again, to validate the interpretation that demonstrability explains the effects.

What, then, is the role of distance in demonstrative reference? In several analyses it was shown that distance modulated whether or not speakers refer demonstratively when pointing by hand, a finding that replicates an earlier result reported by Bangerter (2004). This difference was seen at both coarse- and fine-grained levels. On the coarse-grained

level, Finders used demonstratives more in the NEAR-hand condition of Study #1 than in the FAR-hand condition. Additionally, on the fine-grained level, in the NEAR-hand condition there was a strong correlation between how close a Fribble was to the idealized nose position of the Finder and the rates at demonstratives were used to refer to it. Effects of distance on *which* form of demonstrative is used were somewhat equiovocal. No main effect of distance or any interaction effects were observed in Study #1, though a lack of statistical power may be to blame for these null results. Indeed, the mean proportion of proximal demonstratives was marginally higher in the NEAR-laser (.75) condition compared to the FAR-laser condition (.68)[13]. A *post hoc* analysis of the data from Study #1 revealed that participants in the laser-pointing conditions used a higher proportion of proximal forms for the closest 25% of targets (mean= .74) than for the farthest 25% of targets (mean= .66). Since demonstrability remains constant for all targets in the laser-pointing conditions, this finding suggests that distance does indeed play a role over and above demonstrability. Distance and demonstrability are confounded in the two hand-pointing conditions, so it is impossible to tell what caused the lower rates of proximal usage in the FAR- versus NEAR-hand condition.

In addition to the primary finding that demonstrability plays a role in demonstrative reference, several important secondary findings emerged from these studies. One such finding is that demonstrative use is highly variable across speakers, even in highly constrained tasks in well-controlled physical settings[14]. Part of this

---

[13] Note, however, that the medians showed the opposite pattern (NEAR-condition= .76; FAR-laser condition= .81).

[14] Coventry et al. (2008) also found high variability across speakers (see data tables, pgs. 893-4), though they do not discuss it.

variability may be due to the fact that the particular type of demonstrative reference examined here— namely, exophoric reference to clusters of objects outside of peripersonal space— is relatively unconstrained. Still, the level of variability observed strongly suggests that an account of demonstratives based on antecedent, sheer physical parameters of the context is inadequate. Instead, the data support an alternative view that demonstrative use is irreducibly subjective, a point of emphasis in several theoretical accounts (e.g. Hanks, 2005). Variability is likely not uniformly high in all referential contexts. A priority of future work should be the identification of contexts in which demonstrative use is highly constrained, as well as contexts in which it is less so. Both are informative.

Another noteworthy finding is that bodily action— in particular, pointing— and choice of demonstrative form are very clearly related. Speakers use a higher proportion of proximals when pointing than when not pointing. This pattern appears strong, and is reliable across different task settings and different roles. It was observed for Finders in all four conditions of Study #1, as well as for both Finders and Recorders in Study #2. A similar finding was reported by Piwek et al. (2008) for Dutch demonstratives. In their analysis of 93 initial references to Lego blocks in a controlled referential communication task, they found that speakers used proximals in 26/50 (52%) cases when pointing but only in 1/43 (2%) cases when *not* pointing. This finding suggests that the relationship between pointing and demonstratives is not merely one of incidental co-occurrence. Indeed, the association between pointing gestures and proximal forms may contain a crucial clue to a semantic basis of demonstratives (c.f. Piwek et al., 2008). At the very

least, such findings strike a cautionary note against analyzing demonstrative usage without attention to the co-involvement of bodily action.

While the results of these studies suggest that non-spatial factors are involved in demonstrative reference, they do not uniquely support one particular encompassing framework. Indeed, the data are in principle compatible with an accessibility account (Hanks, 2005), or a control account (Brovold & Grush, *in press*). One emerging theoretical framework that appears especially promising in accounting for the disparate findings reviewed above centers on the notion *deictic force*. *Force*, or *intensity of indicating*, has been proposed previously as a critical parameter in demonstrative reference by several authors (e.g. Wu, 2004), but has only been sketched in bare-bones form (see Piwek et al., [2008] for a recent proposal, but one which differs from the account offered here). The central idea of the deictic force account is that a speaker's choice of demonstrative forms reflects an effort to re-orient an audience's attention more or less intensely. Proximal terms mark high deictic force, or intense indicating. Crucial for such a force-based account is the linguistic notion of *privativity*. Several authors have speculated that the distinction between proximal and distal is actually better conceived of as an asymmetric one, comprising a marked term— i.e. proximal— and a neutral unmarked term— i.e. *non*-proximal (Enfield, 2003; Levinson, 2004). This insight is usefully incorporated into the deictic force account: proximal forms mark high force, while the distal terms mark not low but *neutral* deictic force, an absence of high force. Disparate factors in context might motivate defaulting to a neutral term over the high-force term. If, for instance, the audience is already attending to the referent, a high-force term is pragmatically inappopriate. If the hearer is unlikely to successfully locate the

referent— in other words, if attention re-orientation is insecure— then a high force term may also be inappropriate. Force, it should be emphasized, is ultimately a subjective aspect of the utterance. Speakers' use of deictic force may co-vary in predictable ways with concrete features of context, but it cannot be reduced them.

The deictic force account sketched above remains imprecise in several particulars. Yet even in provisional form it offers a compelling explanation for several of the present findings. First and foremost, it explains the relation between demonstrability and observed usage patterns. Finders are more likely to use high-force terms when they are confident they will successfully reorient audience attention. Second, it accounts for the high levels of inter-participant variability. Degree of force is a subjective feature of an utterance, in stark contrast to distance from the speaker, which is an objective feature of physical context. Third, it explains the striking finding from Study #2 that Finders univocally use the neutral force term— *that*— when referring to a Fribble the Recorder is concurrently pointing to. Pointing toward an entity is an unmistakeable signal that one is also attending to that entity. Thus, when the Finder sees the Recorder pointing to a target, forceful re-orientation would be gratuitous if not confusing. It is unclear what explanation a distance-based account— whether metric, conceptual, or relational— might offer for this particular pattern. Fourth, it explains the association between pointing gestures and proximals. Pointing is an overt bodily manifestation of the effort to reorient attention (c.f. Piwek et al., 2008); it thus makes sense that it would align with the high force demonstrative term. Lastly, and significantly, part of the appeal of the deictic force account is that it explains both exophoric and endophoric uses of demonstratives in one breath by appeal to their common semantic substrate. Reference to previous discourse, or

anaphora, is marked with a neutral force form (*that*) because the referent is presumed to be already in the listener's attention. Reference to upcoming discourse, or cataphora, is marked with a high force form (*this*) because attention is being re-oriented to something new. The deictic force account as just outlined is broadly consistent, then, with a number of recent approaches that seek to ground demonstrative use in attentional dynamics (Burenhult, 2003; Küntay et al., 2006).

Demonstratives have long been of special interest to philosophers, linguists, and anthropologists. Though they have been ignored in the cognitive sciences until very recently, their locus at the intersection of several subfields— including language production and processing, joint action, and co-speech gesture studies— makes them fertile ground for further research. The provisional proposal offered here relates demonstratives to pointing affordances and attentional dynamics, highlighting their intrinsically subjective and multimodal nature. Such apparent complexities make them all the more tantalizing as an interdisciplinary puzzle. Future work on demonstratives stands to illuminate— not just the semantics of a handful of high-frequency lexical items— but the very nature of the cognitive processes underlying everyday reference.

References

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*, 415-419.

Bangerter, A., & Chevalley, E. (2007). Pointing and describing in referential communication: When are pointing gestures used to communicate? In I. van der Sluis, M. Theune, E. Reiter, & E. Krahmer (Eds.), *CTIT Proceedings of the Workshop on Multimodal Output Generation (MOG)*. Aberdeen, Scotland.

Berti, A., & Frassinetti, F. (2000). When Far Becomes Near: Remapping of Space by Tool Use. *Journal of Cognitive Neuroscience*, *12*(3), 415-20.

Brovold, A. & Grush, R. (*in press*). Towards an (improved) interdisciplinary investigation of demonstrative reference.

Bühler, K. (1990 [orig. 1934]). *Theory of Language: The representational function of language*. Philadelphia: John Benjamins.

Burenhult, N. (2003). Attention, accessibility, and the addressee: The case of the Jahai demonstrative *ton*. *Pragmatics*, *13*(3), 363-379.

Clark, E. V., & Sengul, C. J. (1978). Strategies in the acquisition of deixis. *Journal of Child Language*, *5*, 457-475.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 245-258. doi: 10.1016/S0022-5371(83)90189-5.

Clark, H. H., & Krych, M. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62-81. doi: 10.1016/j.jml.2003.08.004.

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near-far perceptual space and spatial demonstratives. *Cognition*, *108*(3), 889-95. doi: 10.1016/j.cognition.2008.06.010.

Deissel, H. (1999). *Demonstratives: Form, function, and grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, *17*(4), 463-489. doi: 10.1515/COG.2006.015.

Dixon, R. M. W. (2003). Demonstratives: A cross-linguistic typology. *Studies in Language*, *27*(1), 61-112.

Enfield, N. J. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, *79*(1), 82-117.

Eriksson, M. (2009). Referring as interaction: On the interplay between linguistic and bodily practices. *Journal of Pragmatics*, *41*(2), 240-262. doi: 10.1016/j.pragma.2008.10.011.

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

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

Hanks, W. F. (2005). Explorations in the Deictic Field. *Current Anthropology*, *46*(2), 191-220. doi: 10.1086/427120.

Hanks, W. F. (2009). Fieldwork on deixis. *Journal of Pragmatics*, *41*(1), 10-24. doi: 10.1016/j.pragma.2008.09.003.

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

Jarbou, S. O. (2010). Accessibility vs. physical proximity: An analysis of exophoric demonstrative practice in Spoken Jordanian Arabic. *Journal of Pragmatics*, *42*(11), 3078-3097. doi: 10.1016/j.pragma.2010.04.014.

Jungbluth, K. (2003). Deictics in the conversational dyad: Findings in Spanish and some cross-linguistic outlines. In F. Lenz (Ed.), *Deictic Conceptualisation of Space, Time and Person* (pp. 13-40). Philadelphia: John Benjamins.

Kemmerer, D. (1999). Near and far in language and perception. *Cognition*, *73*, 35-63.

Küntay, A. C., Özyürek, A., & Planck, M. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *Journal of Child Language*, *33*(02), 303. doi: 10.1017/S0305000906007380.

Levinson, S. C. (2004). Deixis. In L. Horn (Ed.), *The handbook of pragmatics* (pp. 97-121). Oxford: Blackwell.

Lyons, J. (1977). *Semantics (2 vols.)*. New York: Cambridge University Press.

Laury, R. (1996). Conversational use and basic meanings of Finnish demonstratives. In A. Goldberg (Ed.), *Conceptual Structure, Discourse and Language* (pp. 303-319). Stanford, California: CSLI.

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, *8*(2), 79-86. doi: 10.1016/j.tics.2003.12.008.

Miller, G. A. (1982). Some Problems in the Theory of Demonstrative Reference. In R. J. Jarvella & W. Klein (Eds.), *Speech, Place, and Action: Studies in Deixis and Related Topics* (pp. 61-72). London: John Wiley & Sons.

Piwek, P., Beun, R., & Cremers, A. (2008). "Proximal" and "distal" in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, *40*, 694-718. doi: 10.1016/j.pragma.2007.05.001.

Talmy, L. (2000). *Toward a Cognitive Semantics (2 vols.)*. Cambridge, MA: MIT Press.

Wu, Y. (2004). *Spatial Demonstratives in English and Chinese: Text and Cognition*. Philadelphia: John Benjamins.

Chapter 3. Body-directed gestures

3.1. Body-directed gestures

A little more than eight minutes into a 2006 television interview, then-president of the United States George W. Bush pointed to the free world. He and the interviewer, a journalist for the Irish network RTÉ, were discussing the state of the Iraq war. Visibly exasperated by the interviewer's persistent and unwelcome focus on the conflict's death toll, Bush insists: "These people are willing to kill innocent people. They're willing to slaughter innocent people to stop the advance of freedom." He continues: "So the free world has to make a choice." As he says "free world" he produces an apparent pointing gesture, drawing his hand quickly toward his body so that it comes to rest with palm flat on his torso. After a pause, he continues: "Do we cower in the face of terror, or do we lead in the face of terror?" The gesture quietly asserts a connection between Bush and a much broader and somewhat abstract entity— "the free world"— a connection which is then made more explicit seconds later in his use of the pronoun *we*. While such gestures are all the more conspicuous in moments of pitched and public statecraft, they are by no means confined to such contexts. Speakers quite commonly direct gestures at their own bodies: they produce such gestures in all kinds of everyday discourse contexts, and pair them with a wide variety of spoken referents. Perhaps this should not be surprising. The body is an ever-available resource in face-to-face interaction, a rich site for anchoring meaning of all kinds. The present paper considers body-directed gestures produced by English speakers, with particular focus on their relation to co-produced speech, their form and function, and the conceptual processes that motivate them.

Consider the following more mundane and perhaps more commonplace example[15], taken from a television interview in which the topic of discussion is personal rather than international struggle. The interviewee is Natalie Cole, a popular American singer, and she is opening up to the interviewer, Tavis Smiley, about unfulfilling relationships in her past:

Example 1 [NC 11:31]

NC: 1 when I look back on, you know, some of the people that I've known
2 they've all wanted to- they didn't und- they didn't get it
3 they didn't get **me**

As the speaker says "me" in line 3, she quickly brings both hands toward her midline, her fingers loosely bent toward her body such that they approach the center of her torso. She holds them in this position for close to two seconds before continuing. *Me*— and other variants of the first-person singular pronoun, such as *I* and *my*— are often co-produced with body-directed gestures. But certainly not all instances of body-directed gestures are quite so transparent. Presented below are instances of gestures co-produced with first-person *plural* pronouns (e.g. *we*), the second person pronoun (*you*), body-part words (e.g. *face*), abstract notions related to subjective experience (e.g. *comfortable*), and seemingly far-flung concrete nouns (e.g. *script*). Drawing on examples like (1) from a corpus of Tavis Smiley interviews, this paper discusses these and other cases in which speakers

[15] Examples are transcribed to include all spoken words, with each numbered line corresponding to an intonation unit. Segments of the speech co-produced with a gesture of interest are bolded and underlined; other gestures are not marked. Dashes indicate disfluencies in the speech. All clips are identified by the first and last initial of the interviewee, along with the start time of the segment.

produce gestures that indicate, in strikingly different ways, their own bodies. It distinguishes three common types of body-directed gestures. A first type, exemplified by (1) above, involves pointing to the body *qua* person. This type will be termed a *self-point*. A second type involves indicating, either by pointing or touching, a particular body part in the course of direct reference to that part, such as the knee or ear. This type is termed a *body-point*. A third type involves anchoring experiential notions— in some cases quite abstract— to a particular region of the body. This third type will be termed a *body-anchor*. Each type reveals complexities in speakers' spontaneous use of the body as a basic meaning-making resource, as well as in the ways that speech and gesture co-express referential meaning.

Body-directed gestures have gone largely unexamined in the literature on co-speech gesture, aside from a few mentions made in passing. In his nineteenth-century treatise on Neapolitan gestures, Andrea de Jorio (2000 [orig. 1832]) notes that "the palm of the hand close to the chest, whether pressing upon it or touching it gently, just with the tips of the fingers, denotes *me* or *to me*" (pg. 268). De Jorio's example bears a clear resemblance to modern-day self-points among speakers of American English, such as (1), which will be discussed in detail in the next section. Researchers since de Jorio have only occasionally alighted on body-directed gestures as a phenomenon of interest. For example, in her work on the semiotics of French gesture, Calbris (1990) includes an appendix on "gestures targeting the body" (pgs. 222-5). Many of Calbris' examples are highly conventionalized, bearing a clear relationship to idiomatic expressions related to drunkenness, dishonesty, satiety, among others, and often involving rigid standards of form. The present work instead focuses on spontaneous, idiosyncratic co-speech gestures

that have no obvious basis in idiomatic language and, as will be demonstrated, less rigid standards of form. Streeck (2009) discusses a handful of examples of gestures from a psychiatric interview, which include body-directed gestures accompanying notions of "consciousness" (pg. 154) and "feeling" (pg. 157). According to the typology outlined below, such gestures would be considered body-anchors. As Streeck notes, such gestures seem to exploit associations between parts of the body and experiential concepts. Metonymic associations of the sort Streeck appears to have in mind— some highly conventional, others less so— will be discussed in more detail below. Elsewhere, Streeck briefly describes processes of "self-marking" in which speakers index their own bodies in making reference to *others*' physical attributes (e.g. "angel-like hair") or clothing (e.g. "white skirt"). Streeck refers to such examples— somewhat cryptically— as cases of "indexical reference once removed" (pg. 143). Again, similar examples are discussed below in considerable detail. Indexicality appears in many different guises in co-speech gesture, and one aim of the present study is to examine these guises and the conceptual processes underpinning them.

What the different types of body-directed gestures— *self-points*, *body-points,* and *body-anchors*— have in common is that they indicate the speaker's body, either the whole or some particular part of it. Body-directed gestures thus present a diverse microcosm of gestural indexicality. They are not all "pointing gestures", at least not in the prototypical sense in which the term is sometimes used. Indexicality in gesture is sometimes narrowly identified with pointing, and, indeed, many of the gestures analyzed in the present study appear to be what Kendon has called "pure pointing" (1980, pg. 106). Yet it is argued below that even in their purportedly purest of incarnations, pointing

gestures involve complexities of different kinds that go unnoticed (c.f. Haviland, 2003). In the prototypical case[16], a pointing gesture is assumed to involve at least a few key features. First, an index-finger extended handshape is used. Second, whatever is pointed to is concrete, perceptible, and present in the speech situation. Third, if the pointing gesture is co-produced with speech, it is taken for granted that what is pointed to— the *demonstratum*, or *target*— is in some sense identical to what is simultaneously referred to in speech— the *denotatum*, or *referent* (c.f. Clark et al. 1983). As an example of such a prototypical pointing gesture, think of a customer approaching a display case in a donut shop. When it comes time to order, the customer says "I'll have a cruller" while indicating one from a row of crullers with an extended index finger. Canonical morphology is used, the referent is perceptually available, and the target and referent form an unproblematic match. Co-speech body-directed gestures, in contrast, violate one or more of these prototypical features in just about every instance. For starters, self-points only rarely involve an index finger extended morphology, more often employing a full hand, as in (1). Further, speakers find occasion to point to entities that are not concrete and perceptible, but abstract and experiential, such as *instinct* and *belief*. Part of the theoretical interest of body-directed gestures lies precisely in their curious non-conformity to the prototype.

Of considerable interest is the fact that, in body-directed gestures, the relation between what is pointed to— hereafter, the *target*— and what is said— hereafter, the

[16] The notion that there is a prototypical pointing gesture, or scenario, has been suggested in the cognitive linguistics literature. Langacker (2008) has described the "canonical pointing gesture" as a "conceptual archetype" (pg. 284), noting that it is often enshrined in directional signage. Lakoff (1987, pg. 490-1) describes the "Pointing-Out [Idealized Cognitive Model]".

referent— is often not altogether straightforward. This referential complexity is evident, almost comically so, in Bush's "free world" point above. Clark et al. (1983) have discussed a hypothetical example to illustrate this same phenomenon, in which a woman says "I used to work for those people" while pointing to a copy of the *New York Times*. In this example, the target— a physical copy of a newspaper— and the referent— the owners of the newspaper— are hardly identical, but instead stand in an apparent metonymic relation of product for producer. In more recent work, Clark (2003) has introduced the concept of a "chain of indicating" (pg. 264) to characterize such cases: the speaker indicates the *demonstratum*, which in turn indexes the referent. *Target-referent* metonymy of this sort may be a common feature of pointing in real-world settings, though it has been described only spottily. People point to tokens to refer to types, to objects to refer to their functions, to parts to refer to wholes, to present persons to refer to their non-present affiliates, and so on (c.f. Haviland, 2000, pg. 32; see also Morford & Goldin-Meadow, 1997, pg. 431 on "non-literal pointing" among homesign users). Below I argue that such metonymies— where they arise in co-speech gesture— are motivated by general conceptual processes of compression (Fauconnier & Turner, 2002).

But *target-referent* metonymy is only one kind of metonymic relation evident in pointing gestures. Imagine now a scenario in which someone points to a wall and says "This wall is going to be painted yellow". Arguably, to fully understand the utterance one must recognize the relation between the particular part of the wall that is pointed to and a larger entity, the entire wall. It will thus be useful to further distinguish between the *site*— that is, the actual spot where a pointing gesture makes "contact" with the material world— and the *target*— in many cases a larger entity that encompasses the *site*. *Site-*

target metonymy is also a ubiquitous feature of pointing, yet such metonymies are often hard to see precisely because they are so commonplace and automatically interpreted. Example (1), as well as numerous other torso-directed gestures co-produced with personal pronouns, draws on a site-target metonymy whereby the center of the chest stands for the entire person. In fact, de Jorio noticed this particular part-for-whole metonymy and ventured that "the reason why indicating one's own chest serves to denote one's own being, the *me*, is because in doing this one indicates the heart, which is the seat of human life, and hence one's being" (pg. 269). More important than the plausibility of de Jorio's particular proposal is his insight that the interrelations between *site*, *target*, and *referent* cannot be taken from granted.

Another complexity of pointing gestures is that they are often leavened with iconic properties. Kendon (2004) draws a distinction between pure pointing— in which little more than pointing is done— and other kinds of pointing in which referents are simultaneously pointed to and *characterized*. To illustrate this latter impure kind of pointing, he describes a case in which a woman in an Italian marketplace indicates a vegetable scale while wagging her finger back and forth. The gesture indicates the location of the object in the world at the same time that it iconically characterizes the motion of the needle on the scale (Kendon, 2004, pg. 202). Similar examples of indexical-iconic hybrids have been discussed elsewhere in the ethnographic literature on pointing (e.g. Haviland, 2003, pg. 151). Hybridity in pointing gestures is a matter of degree. As I argue below, subtle construal operations may be at play in even the most innocuous morphological alternations, such as between an extended index finger and a flat hand.

Much as pointing is taken to involve a suite of prototypical features, it is often taken to perform certain prototypical functions. Foremost is the function of re-orienting a listener's attention to some object in the world. Charles S. Peirce (1940) attributed this function— namely, that "they direct attention to their objects by blind compulsion" (pg. 108)— to indices generally, a broad category in which he includes weathervanes and thunderbolts as well as pointing fingers. Attention re-orientation is the function of pointing widely taken for granted not only in contemporary research on co-speech gesture (e.g. Clark, 2003), but also in most current research on pointing in animals (e.g. Xitco et al., 2004), in child development (Tomasello, et al., 2007), and in sign language (Pizzuto & Capobianco, 2008). Indeed, the prototypical pointing scenario as sketched above is taken by some commentators to be a paradigm case of joint attention generally (e.g. Kockelman, 2005).

A further function often attributed to pointing is that it supports precise, unambiguous reference (c.f. Bangerter, 2004). Pointing does not merely shift attention, it shifts attention to a unique referent[17], and it is co-designed with speech to do so unambiguously. Notice that, on the face of it, cases of self-pointing seem to violate one or both of these functional assumptions. When the speaker in (1) points to herself, she is not *re-directing* the interviewer's visual attention; he is already looking at her. Nor does the gesture support a reference that would otherwise be ambiguous: first person pronouns can hardly be claimed to require gestural disambiguation. To the extent that body-directed gestures often confound the commonplace view of why we point— at least in

[17] Peirce hints at this function when he writes that a defining characteristic of indices is that they "refer to individuals, single units, single collections of units, or single continua" (Peirce, 1940, pg. 108).

caricature— they present a vital opportunity to sharpen up accounts of the function(s) of gestural pointing as a broad, heterogeneous class.

This paper has several interrelated aims. Given that body-directed gestures have largely escaped empirical attention, the first aim is fundamentally descriptive. Two basic descriptive questions will be considered in detail: *when* do speakers point to their bodies, i.e. with what kinds of co-produced speech and in what discursive contexts; and *how* do speakers point to their bodies, i.e. with what morphologies and movement dynamics. By hypothesis, the questions can also be productively fused: How are particular morphologies related to particular contextual factors? If such a relationship exists, at what level of specificity does it hold? A second, more theoretical aim is to venture steps toward an answer to a deeper question: Why do speakers point to their bodies? What conceptual processes motivate such gestures and what discursive circumstances prompt them? All of these questions— *when*, *how*, and *why*— ramify far beyond the microscopic case study of body-directed gestures considered here, to an understanding of co-speech gesture generally.

3.2. Data and analysis

The data for this study are a corpus of one-on-one interviews from the Tavis Smiley Show, an American television program that airs nightly on PBS. The Tavis Smiley Corpus (hereafter, abbreviated as the TSC) consists of forty interviews, totaling more than eight hours of conversation, with each interview ranging in length from 9 to 24

minutes[18]. The host of the show, Tavis Smiley, is a 47 year-old man; the interviewees are twenty women and twenty men, ranging in age from 29 to 77 with a mean age of 49.44 years old[19]. All interviewees are native speakers of American (or, in one case, Canadian) English. The interviewees are singers, television and film actors, authors, directors, activists, and so on. Politicians were deliberately excluded from the corpus on the grounds that they are often deliberately coached on both the frequency and style of their gesture production. Certain features of the Tavis Smiley Show make it an especially good trove of freely available, unscripted gesture data. For one, in contrast to other televised interview programs, the camera angles invariably afford a full view of the speaker's gesture space. Second, there is no live studio audience, and aside from extremely rare banter between the host and the on-site production team, the conversations are dyadic. Third, the interviews cover a wide range of topics, including professional triumphs and failings, experiences with illness, and the creative process, among many others[20].

Each interview was analyzed for instances of body-directed gestures and was annotated using ELAN video annotation software (available online: http://www.lat-

---

[18] Note that the interviews are made up mostly, but not entirely, of spontaneous dialogue. Each video includes a generic introduction to the show, followed by a rehearsed introduction of the guest read aloud by Smiley. Further, the dialogue is occasionally interrupted by other media, such as a clip from a performance by the interviewee.

[19] The mean does not include one interviewee whose age could not be readily determined from online sources.

[20] Data from the TSC is not without limitations. One clear limitation is that the camera often only shows the current speaker, not the addressee. This makes certain kinds of analyses, for example, about the possible role of audience behavior during body-directed gestures difficult to address. Another limitation is that, while there is no studio audience, there is an implicit viewing audience— those watching the program on television. A final limitation is that, though every effort was made during corpus construction to include as broad a range of backgrounds as possible, many of the interviewees are performers, a fact which reflects the composition of the show's guests as a whole.

mpi.eu/tools/elan/). On a final pass through the corpus, the following features of the body-directed gestures were coded for: co-produced speech, including precisely co-timed speech as well as larger utterance context; handshape; handedness of gesture (right, left, or bimanual); in cases of bimanual gesture, whether the motion of the hands was symmetric or asymmetric; what part of the body was targeted; and, finally, whether or not contact was made between the hands and the body.

Determining the presence of a body-directed gesture was not always straightforward, for a few reasons. One reason is that body-directed gestures are frequently realized as relatively small movements, both in terms of intensity and amplitude. Further, the video angles available in the TSC are most often front-on— along with occasional 45 degree side angles— and thus do not afford a clear view of movements along the speaker's sagittal body axis. These difficulties are compounded by the fact the some speakers make frequent beat-like gestures along the sagittal axis. Reliability was assessed by having two additional coders analyze a subset of the TSC (10 interviews, or 25%) for the presence of body-directed gestures. The two coders annotated the videos separately and then resolved any disagreements collaboratively. After agreeing on a final set, their annotations were compared to those of the author. 84% (107/127) of the body-directed gestures annotated by the author within the subset were also annotated by the additional team of coders[21]. All statistics presented below are based on the final coding of the complete data by the author. Beyond video coding for the particular

[21] However, the additional coders also annotated a number of instances that where were not included in the author's analysis. Overall only 64% (107/167) of the gestures annotated by the collaborative team were also annotated by the author. The author was thus considerably more conservative than the coders in the determination that a given movement counted as a body-directed gesture.

features identified above, additional analyses were conducted to address specific questions, and these are described in more detail below as relevant to the discussion. Where appropriate word-by-word transcripts of each interview were retrieved from the Tavis Smiley Show's web archives (www.pbs.org/wnet/tavissmiley/archive/) and were used to determine token counts for certain words of interest.

3.3. Self-points

The first category of body-directed gestures, and also the most commonplace, is that of self-points. In a self-point, a speaker points to some part of the body— in the present data, invariably the torso— to evoke, as it were, the whole self. Self-points in the TSC are usually but not always co-produced with person terms. Of the 207 self-points (55 produced by Tavis; 152 produced by his guests), the majority (n=156 [75.4%]) were co-produced with the first-person singular pronoun and its derivatives: *I* (n= 85), *my* (n= 33), *me* (n= 29), *myself* (n= 7), and *mine* (n=2). Consider the following example, from an interview with the writer and actor Hill Harper, in which Tavis Smiley discusses his own misgivings about writing a memoir:

Example 2 [HH 09:43]

TS: 1 I've got so much life I wanna live
2 what business do **I** have trying to write a story about **my** life?
3 I finally came around to doing it very quickly

Figure 3.1: Three frames from example 2. Tavis produces two self-points, first with "I" (a) and then with "my" (c).

Smiley points to himself twice in succession, first with "I" (line 2) and then with "my" (line 2) (see Figure 3.1). In both cases the right hand is held loose and is drawn toward the middle of the torso during the stroke phase of the gesture. First-person pronoun terms are extremely common in English. In the Corpus of Contemporary American English (COCA), collapsing across all discourse types, *I* is the 11th most common word, *my* is 44th, and *me* is 61st. In the TSC, *I* is the 4th most common word— ranking after only *and*, *of*, and *to*— and is thus the most common noun[22]. Other first-person pronoun words are somewhat less common, but still rank in the hundred most frequent words (*me*= 29th , *my*= 31th). Curiously, self-points accompany these words relatively infrequently. *I* is co-produced with a self-point in a mere 2.34% of instances (85/3625); *my*= 6.55% (33/504); *me*= 5.56% (29/522); *myself*= 12.96% (7/54)[23]. As is evident from these percentages, first-person pronominal words are not equally likely to be accompanied by a self-point (χ^2= 51.54, df= 3, p< 0.001). Such low percentages invite several questions. What, if

[22] Frequency counts were obtained by text analysis of the word-by-word transcripts provided on the program's website. The corpus includes 102,781 words.

[23] These should be considered conservative— and slightly noisy— rates. Speakers are not always visible on camera, and, moreover, though the cameras tend to track the primary speaker, the margins of utterances are occasionally cut-off. There are probably some instances of self-points happening off-camera.

anything, is special about those relatively rare instances in which self-points *do* crop up? Is the presence of a self-point relatable to the *case* of the pronoun—subject, object, possessive, reflexive— or perhaps to the broader discourse structure? These questions are taken up in more detail below.

Somewhat surprisingly, the corpus also contains a number of instances of self-points co-produced with the first person plural: *we* (n= 12) and its grammatical affiliates, *us* (n= 7) and *our* (n= 7). In the following example, Tavis is discussing the complicity of various parties— congress, the media, and the public— in the ongoing conflict in Iraq:

Example 3 [PD 07:10]

TS:	1	as American people we know that this thing is-
	2	has gone, has gone awry
	3	and how much are **we** to blame for allow-
	4	I thought it was we the people
	5	I mean how much are **we** to blame for allowing this to go forth still

In line 3 Smiley produces a two-handed self-point, and he repeats the gesture in more or less identical form in line 5. In both cases, the fingers are adducted and drawn rapidly toward the middle of the torso. Self-points co-produced with plural person terms are *prima facie* puzzling. After all, what is referred to in speech— "we"— and what is pointed to in space— a location on Smiley's chest— are not the same; they appear to stand in a metonymic relation of part for whole. Such instances provide a straightforward and minimal example of gestural metonymy, a phenomenon that is both pervasive and variegated in body-directed gestures.

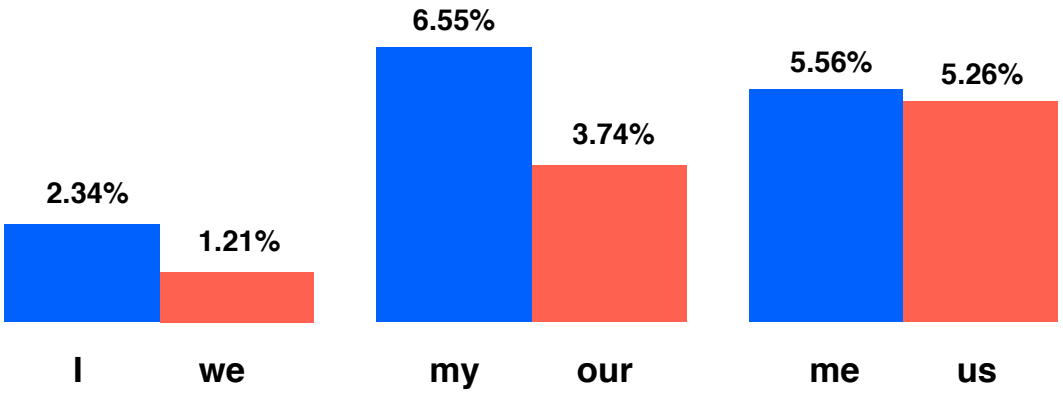

Figure 3.2: The rates at which different pronouns are accompanied by self-points in the Tavis Smiley Corpus. Rates of other pronominal words— *mine, ours, myself*, and *ourselves*— are not included due to insufficient data.

Interestingly, the rates of pointing across pronoun cases pattern in a similar way for both first-person singular and plural variants (see Figure 3.2). *We* (subject pronoun) is co-produced with a pointing gesture in 1.21% of cases (12/984); *our* (possessive pronoun) in 3.74% of cases (7/187); and *us* (object pronoun) in 5.26% of cases (7/133). Like their singular counterparts, first-person plural words are thus not equally like to be co-produced with a self-point; and like their singular counterparts are produced at different rates according to case (χ^2= 13.22, df= 2, p= 0.001). This observed co-patterning lends weight to an inference that the different pronoun cases— subject, object, and possessive— are differentially associated with self-points. The interesting question is why. One possibility is that higher-order factors— such as information structure or grammatical construction type— are the real source of the observed pattern, with pronoun case merely an imperfect correlate.

On rare occasions speakers co-produce self-points with the second-person pronoun *you* and its affiliates (n= 9). In the next example, the actor Mark-Paul Gosselaar discusses his new television series and its heady themes:

Example 4 [MG 04:15]

MG: 1 on our show you'll say
2 really?
3 it'll- it'll make **you sort of question your own beliefs**
4 **and your own** morale

Beginning with "you" in line 3, the speaker draws both hands toward his midline and together so that they touch at the knuckles; with "question your own" (line 3) he brings the hands still closer to his body in a second stroke. In this example, as well as others in the TSC involving second-person pronouns, broader contextual factors indicate that the intended sense of *you* is not the addressee— Tavis Smiley— but the impersonal, or generic *you* (see Whitley, 1978). Of course, given that the impersonal *you* is relatively common in speech, but only very rarely accompanied by a self-point, it is worth asking whether other factors might be at play in this example. Notably here the pointing gesture occurs with a statement about a reflexive, self-directed action— namely, the act of "question[ing] your own beliefs". Reflexivity is foregrounded lexically in the use of "own", which is repeated in line 4. Examples like this of self-points co-produced with second person pronouns have been discussed by McClave (2000) as "pragmatic devices for indicating participant roles" (pg. 248). Specifically, according to McClave's analysis, in usual circumstances the use of the second-person pronoun accompanied by direct gaze

Figure 3.3: Three frames from example 5. Thurston Clarke produces a two-handed self-point while saying "himself" (b).

at the listener strongly suggests a personal reading of the pronoun *you*/ *your* as referring to the hearer. A speaker might use a self-point in such cases in order to disambiguate between an expected personal and the desired *im*personal reading. Unfortunately, much more data would likely be required to adjudicate between these two possible functional explanations— the foregrounding of reflexivity and the disambiguation of personal/ impersonal senses of *you*.

A handful of times in the corpus, speakers point to themselves along with third person pronouns such as *her*, *himself*, or *themselves*. In (5) the historian Thurston Clarke is discussing some early stumbles in Barack Obama's 2008 presidential campaign:

Example 5 [TC 03:53]

TC: 1 I think Barack Obama is held back
2 not by things **within himself**
3 but by the American people

In such cases, self-points seem to supply non-verbal evidence of radical viewpoint shifts of the sort that have been analyzed extensively in the cognitive linguistics literature on

narrative (e.g. Dancygier, 2008). But it is not clear a whole-hog viewpoint transposition needs to be posited to explain (5). Much as it was in example (4), reflexivity is central here and is made explicit in the use of the reflexive pronoun form "himself" (line 2). This lends support to the proposal that self-points may be found wherever reflexivity is foregrounded in the discourse, regardless of whether the speaker is the same as the subject of the reflexive pronoun. Foregrounded reflexivity may also motivate the observed high rate (12.96%) of self-points co-produced with the first-person pronoun *myself*.

How do speakers point to themselves? What forms are used and what factors motivate the use of one form or another? At least two morphological features of self-points set them apart from prototypical pointing gestures[24]. First, self-points are much more commonly produced with some version of a full-hand morphology than with the canonical index finger extended form (see Figure 3.4). Of the 207 tokens of self-points in the TSC, only 20 (or 9.7%) were produced with an index-finger extended morphology; the rest were produced with variants of a loose full-hand morphology (170/207 tokens, or 82.1%) or, in rarer cases, with some other articulated handshape (17/207 or 8.2%), such as a fist (n= 1) or a thumb (n= 2). Why are self-points produced with a full-hand while other kinds of pointing gestures are— at least prototypically— produced with an index finger? One explanation is that pointing morphology embodies the speaker's construal of what is pointed to. Index finger extended handshapes may be good for pointing to focal

[24] One aspect of pointing morphology not discussed here is the length of time that a self-point is held in place. Associations were sought between the duration of the apical hold of the pointing gesture and other factors, such as whether the gesture was co-produced with contrastive stress, but none were detected.

targets like a donut in a display case, but less appropriate for pointing to larger objects, such as persons at close proximity. Pointing morphology on this account is not just a matter of antecedent physical properties of the objects, but of the speaker's subjective, context-driven construal of those objects, a point which will be returned to in the discussion. A second possibility is that, for kinematic reasons, it is awkward or uncomfortable to extend the index finger while adducting the wrist toward the body. A final explanation, perfectly compatible with either or both of the previous explanations, is that full-handed morphology is a conventionalized aspect of pointing to the self among English speakers of North America. To the extent that such a convention exists, however, it is far from rigidly defined. Within the class of full-handed gestures there is considerable variation along a number of dimensions. Sometimes the fingers are spread flat against the chest and other times curled in; sometimes they are splayed wide and other times tightly bunched together.

A further morphological question is whether differences in self-pointing handshape are associated with grammatical characteristics of the co-produced pronouns, such as person (e.g. first or second), number (singular or plural), or case (e.g. subject or possessive). Established sign languages conventionally mark such pronominal differences in many cases with morphological contrasts. For instance, in American Sign Language (ASL) the difference between the subject (*I*) and possessive (*my*) pronoun is marked by the use of index-finger extended and flat-hand morphologies, respectively (Friedman, 1975). It is possible that such differences are not entirely arbitrary conventions but are

instead motivated in some way, their synchronic status as lexemes notwithstanding[25]. If this were the case, similar patterning may be evident in spontaneous co-speech self-points. In the present corpus, no associations were detected between handshape and either pronoun person or number. Curiously, however, handshape and pronoun *case* do not appear to be strictly independent. Though punctate morphologies— collapsing over both index-finger (n= 20) and thumb-extended forms (n= 2)— amount to only 10.6% (22/207) of all cases of self-points, they were not equally distributed across the different pronoun cases. Considering just the first-person singular and plural pronominal self points (n=176), punctate forms are most common among the object pronouns *me* and *us* (8/38), least common among the possessive pronouns *my* and *our* (1/40), and intermediately common among the subject pronouns *I* and *we* (11/98). The difference between these proportions is significant (Fischer's exact probability, n= 176, p= 0.033). The observed pattern echoes— albeit faintly— the ASL pattern in that punctate morphologies appear to be *dis*preferred for possessives. It is also consistent with recent and somewhat informal observations made by Kendon (2010). He writes:

> It seems that, if the speaker is talking about himself as the *agent* of an action or process or as the *source* of an opinion or feeling, he tends to direct a 'flat' hand— all fingers extended adducted— towards his upper chest, sometimes making contact, pressing the palm of the hand against the chest, or sometimes using the side of the hand. In contrast, if he is talking about himself as the recipient of something that comes from outside, he may point to himself by directing an index finger toward his own chest, but not making contact. (pg. 21)

To the extent that Kendon's notion of the *agent* and *recipient* map onto the subject and object pronouns, respectively, his proposed pattern appears to hold. It must be cautioned,

[25] Also relevant to the question of the motivatedness of pronominal morphology in sign languages, of course, would be typological studies of pronominal systems in unrelated sign languages. I am not aware of any such studies.

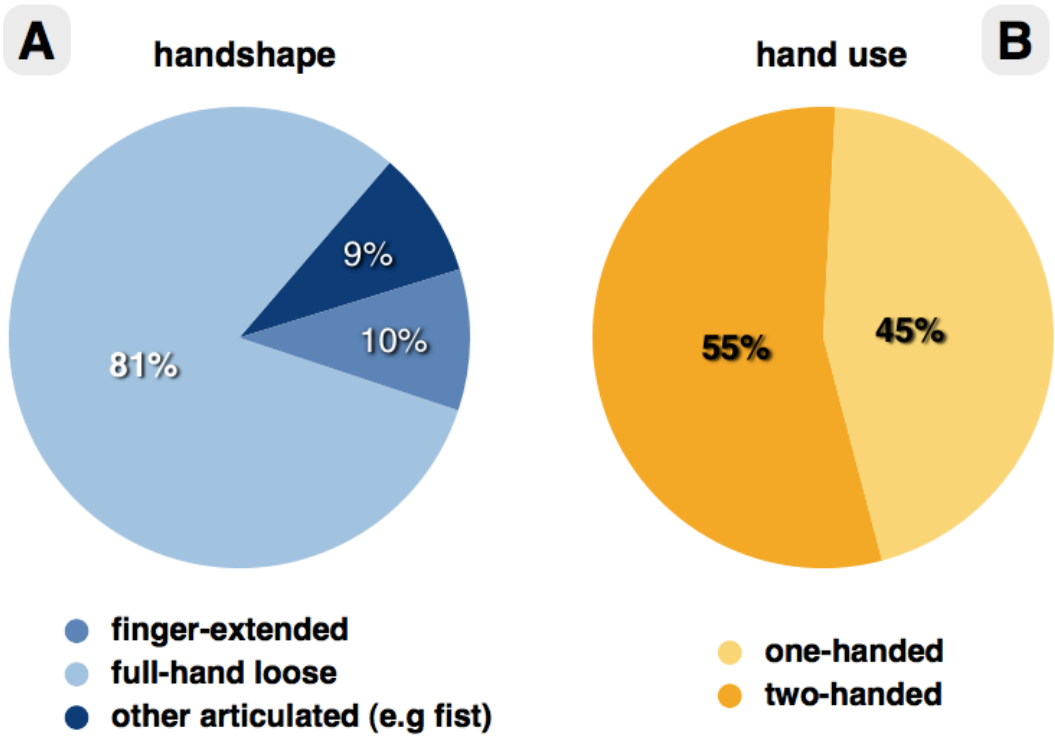

Figure 3.4: Summary statistics of two parameters of morphological variation in self-points. (a) Percentages of different handshapes. (b) Use of one-handed versus two-handed forms.

however, that in the present data any such associations are weak and are clearly not legislated in a strict way. Larger corpora will be required to validate this pattern.

A second distinctive morphological feature of self-points is that they are often bimanual, whereas other types of pointing gestures are only rarely so. Across all 207 cases of self-points 113, or 54.6%, were bimanual (see Figure 3.4). Interestingly, there was a very clear association between bimanuality and handshape, in that bimanual gestures were much less likely than one-handed gestures to involve index finger extension (bimanual= 1/113, one-handed= 19/94, χ^2= 21.96, df= 1, p< 0.001). No associations were evident between bimanuality and grammatical features of the co-

produced pronouns (case, number, person). What factors, then, determine whether a self-point is produced with one or two hands? Does the choice perhaps reflect construal operations? A tempting possibility is that the choice of one- or two-handed variants is driven by graded aspects of discourse context. Indeed, in his very brief discussion of self-points, de Jorio speculated as much, commenting that a two-handed version of the *me* gesture carries the same meaning as the one-handed version "but with greater emphasis" (de Jorio, 2000, pg. 268). Perhaps all pointing gestures can be productively analyzed on a cline of *robustness* (c.f. Fowler & Levy [2000] on "energy peaks" [pg. 215]). In the case of outward-directed points, morphological features determining the degree of robustness include the extension of the arm, the co-oriention of eye gaze and head direction, and the duration of the gesture's apical hold. Such features are a visible manifestation of speaker effort, and as such they may parallel information structure. In support of just such a hypothesis, Enfield et al. (2007) have described a difference between "B-points" ("B" for big) and "s-points" ("s" for small) in Lao pointing gestures, arguing that B-points are used for foregrounded information and s-points for backgrounded but still relevant information. Self-points, by hypothesis, might be morphologically graded in a similar fashion. To test this idea, associations were sought between bimanuality and lexical stress on the co-produced pronoun, on the assumption that if bimanuality is associated with emphasis it might be associated too with *prosodic* markers of emphasis. However, no evidence for such associations was found. It remains possible, of course, that self-points are graded on intensity in ways that are not captured by this analysis, or by other the current coding scheme.

But is bimanuality an entirely arbitrary feature of self-points? A further possibility

is that the use of one or two hands is driven in part by whether the *previous* gesture was one- or two-handed. In other words, the choice of one or two hands in gesture may be subject to effects of what might be termed *morphological momentum*. This idea was investigated in the TSC by analyzing each pronominal self-point for whether it was preceded immediately by rest or else by another gesture[26]. If preceded by a previous gesture, that previous gesture was analyzed for whether it was produced with one or two hands. Of those self-points preceded by *rest*, 68% (or 28/41) were bimanual. Of those self-points preceded immediately by a *bimanual gesture*, 86% (or 57/66) were bimanual. Of those self-points preceded immediately by a *one-handed gesture*, however, only 18% (or 10/57) were bimanual. The difference between these proportions is highly significant (χ^2= 61.85, df= 2, p< 0.001). Use of one or two hands in self-points thus appears to be shaped considerably by effects of morphological momentum.

In terms of their place of articulation, self-points are invariably directed at the torso, usually the upper half of the torso near the mid-line. But precisely where on the torso self-points are articulated is highly variable (see Figure 3.5). In bimanual self-points the hands' place of articulation— in addition to other morphological features— is consistently symmetrical. We might step back and ask why it is, in the first place, that self-points are directed at the chest. Unfortunately, scant data are available on the extent to which this pattern is universal in co-speech gesture cross-culturally. Anecdotal evidence supplies at least one counter-example: Japanese speakers sometimes point to the self by directing an index finger at the nose rather than the chest, a form that is also

[26] Some self-points were excluded from this analysis. Self-points produced by Tavis during which he was holding a prop, such as a book or a CD, were excluded, as were any self-points immediately preceded by other self-points.

Figure 3.5: An illustration of the approximate places of articulation for Tavis Smiley's pronominal self-points.

found in Japanese Sign Language in alternation with a chest-directed point (Yoneyama, 1997)[27]. Possible motivations for pointing to the torso include that the torso is conceptualized as the center of the body's mass or spatial extent, or, more mundanely, that it is easily accessible to the forelimbs from neutral gesture space. De Jorio's suggestion that pointing to the chest is really pointing to the heart as a metonym for the whole person, is fanciful and, ultimately, unlikely. Not all self-points are directed at the position of the heart, and two-handed self-points very often miss the heart on either side.

From a cognitive and functional perspective, perhaps the most interesting

[27] I thank Mike Morgen and Connie de Vos for bringing this reference to my attention.

question we can pose is *why* speakers point to the self. Why does *I* sometimes merit a self-point but much more often not? First-person pronouns are extremely high-frequency words in English, and yet they are only co-produced with self-points in a tiny fraction of instances. They do not serve to disambiguate the referent of an otherwise ambiguous referring expression— by all accounts the voice already does that (c.f. Clark, 2005, pg. 513). Nor do they serve to re-direct the audience's visual attention. This question resonates with a much larger question in gesture studies: Why is gesture present in some utterances but not in others? What determines, within the stretch an utterance, which parts if any will be accompanied by a gesture and which will not? The question has prompted surprisingly little discussion. McNeill (2005, pgs. 54-55) has sketched a notion of "communicative dynamism", proposing that gestures emerge when the discontinuity between prior and present discourse is high. Other attempts at answering the question appear to have ignored pointing altogether (e.g. Hostetter & Alibali, 2008), perhaps on the idea that the reasons people point are self-evident.

Self-points present a highly circumscribed, everyday discourse context in which to explore possible answers to the *why* question, including potential contributions made by discourse context and information structure. The fact that pointing rates are modulated by pronoun case— with subject pronouns less likely to be accompanied by self-points than either object or possessive pronouns— may provide only a small part of the explanation. An important additional clue is found in qualities of the speech stream: several of the self-points considered so far have been accompanied by contrastive stress on the co-produced pronoun (examples 1-3). A full analysis of all of the pronominal self-

points in the corpus revealed that 46% were co-produced with lexical stress[28]. Consider the following example, in which the actor Will Smith talks about the direction of his career and his manager's pronounced role in steering it:

Example 6 [WS 11:58]

WS: 1 uh 'I am Legend' was really the first movie
2 that **I** chose
3 it's always been Jay L

Here the speaker contrasts himself with another contextually available subject who could be doing the "choosing", his business partner and manager, Jay L. Examples already discussed above are amenable to a similar analysis. In (3) the speaker, through the combination of a self-point and lexical stress, twice asserts a contrast between "we" and other possible parties, such as congress and the media. In example (2) the speaker produces six first person pronouns. Yet only two of these are co-produced with self-points, and these are the same two that are co-produced with focal stress. In such cases self-points partner with prosody to disambiguate between a set of possible persons, who— while not necessarily physically present— are contextually available in the discourse (see Zwets, 2009 for a similar observation about the distribution of self-points). On one reading, then, self-points still serve a kind of attention-orienting function, but

[28] For the stress analysis, audio-only clips were prepared of all intonation units containing pronominal self-points in the TSC. Two coders— the present author and a second coder unaware of the purpose of the analysis— analyzed the full set of 334 first-person pronouns occurring in the audio clips, judging whether or not they involved lexical stress. Coders agreed in 87% of cases (Cohen's Kappa= .71, substantial agreement). Only those cases in which both analysts agreed on the presence of stress are included in the 46%.

they serve to orient *discourse* attention rather than visuo-spatial attention. Whether other kinds of pointing gestures— indeed, other kinds of gestures more generally— also embody information structure in this way is promising question for future work (c.f. Enfield, 2009, pg. 79).

Finally, not all self-points are co-produced with personal pronouns. They are also co-produced with certain adverbs (e.g. “personally”) and appositive phrases (e.g. "as a cancer survivor"). In two cases in the corpus, self-points are co-produced with proper names other than the speaker’s. In both cases, an actor is discussing a recently played role and points to himself while saying the name of the character (e.g. "Jerry"). Speakers also point to themselves when referring in speech, not to people at all, but to far-flung referents. In one such case, a speaker, having previously mentioned a script that he co-authored, points to himself while referring back to the "the script". In another case, a speaker who had previously mentioned his upbringing in New Orleans, points to himself while saying "the wetlands". These cases constitute on-the-fly metonymies in which, to take the second example, one part of a frame (i.e. the speaker) is associated with another part of a frame (i.e. the speaker's place of birth). Though considerably rarer in the TSC, they are not different in kind from self-points produced with *we* and *our*, or from Bush’s self-point with “the free world”. A candidate conceptual process at work in such cases— namely, compression— is discussed in more detail in the discussion.

3.4. Body-points

Self-points are directed at a part of the body but evoke the person as a whole. Body-points, by contrast, evoke only a particular part or region of the body. Body-points

are thus the most deceptively simple of the types of the body-directed gestures, and the ones that most resemble garden-variety pointing gestures. Most commonly, in body-points speakers point to or touch a body part while referring in speech to that same body part, to their knees, or ears, or chests. The speech accompanying such gestures includes partonymic terms as well as demonstratives. Body-points are at the same time less frequent (n= 45) and considerably more varied than self-points, in terms of the morphologies and referents involved, and so meaningful quantitative comparisons are difficult to make in the present data. Instead the focus here is on general, qualitative features of body-points that can be observed across disparate examples.

In a first example, Holland Taylor is discussing early-career setbacks and the emotional toll they can take:

Example 7 — HT 02:40

HT: 1 and I had been crushed by not getting it
2 and I had suffered over it
3 and often remembered it when **my face would get hot** with
4 embarrassment

As she says “my face” (line 3), the speaker brings both hands held flat, fingers adducted to her cheekbones. With “get hot” (line 3) she produces a small amplitude outward stroke. Idiosyncratic morphologies and places of articulation are a signature of body-points, and index-finger extended morphologies are somewhat rare (7/45). In (7) and many other cases, highly context-dependent forms are used in ways that appear motivated by spatial properties of the referent, such its extent and topology. For this reason, pointing morphology in all body-directed gestures, but most conspicuously in body-

points, is best understood as a process of *construal*. In (7) the use of two flat hands is probably motivated by a construal of the face as a symmetrically extended surface; the outward stroke, meanwhile, construes the experience of "getting hot" as a sudden one.

Perhaps more interesting are cases in which speakers point to their bodies while referring in speech to a generic human body[29]. The jazz critic Ben Ratliff, in the next example, discusses a recently completed project in which he structured interviews around shared listening sessions with musicians:

Example 8— BR 09:57

BR: 1 and as it was going on we would follow it together
2 um so it was just about you know connecting with people I respect and am interested in
3 **through the ear**

As he begins to say "through" (line 3) he directs the index finger of his left hand upward toward the left side of his head. With "the ear" the finger moves outward from his body before both hands return to rest. In speech Ratliff uses the definite article *the* to make type-level, generic reference to "the ear", yet at the same time the gesture grounds the reference to his own body. Such a distribution of referential labor is common in body-points, with verbal types often anchored to physical tokens on the speaker's body.

Just as generic or non-specific human bodies are invoked in body-directed gestures, so too are the bodies of specific others. In (9) Judy Reyes produces a series of

[29] Floyd (2010) discusses examples of "reflexive gestures" used in discourse about social categories, such as instances of people indicating their own heads to characterize the hair types of other ethnic groups.

body-points while commenting on the striking physical likeness between a particular mother-daughter pair:

Example 9— JR 03:01

JR: 1 the girl is identical to her mother
2 from the- the **face to the mouth** to the **hairdo** even
3 they both have **bangs** and a **ponytail**

Reyes produces four body-points in quick succession, each with a different handshape and movement dynamics. As she says "face" (line 2) she brings both hands, fingers spread, to just below her jaw and holds them while continuing to say "to the mouth"; with "hairdo" she sweeps both hands backward and slightly downward before returning them to rest. In line 3, as she says "bangs and ponytail" she sweeps her left hand over her forehead while simultaneously bringing her right hand to a position immediately behind her head. Of particular interest are the movement patterns involved in the gestures co-produced with "bangs" and "hairdo". The gestures do not merely point to positions on Reyes' body; they characterize them has having a particular spatial extent and downward motion (though the latter is fictive in Talmy's [2000] sense).

Body-points such as those in (9) may appear entirely pedestrian, but they hide certain complexities. For instance, in virtue of what conceptual processes can a speaker point to her *own* body in the course of reference to another's bodily features— features that, moreover, she does not herself have? What makes such a gesture readily intelligible rather than outright confusing? In (9) and related examples, it is difficult to adjudicate between several possible analyses. According to one analysis, the speaker is shifting

viewpoints to the body in question— that is, it is as if she is pointing to her own body from another's perspective. Such an analysis is plausible: viewpoint shifts are a pervasive feature of gesture, and in some cases multiple viewpoints are simultaneously evident in a gesture (see Parrill, 2009; also McNeill, 1992 on "chimeras"). A different analysis might posit that the speaker is using the body as a kind of ready-made representational structure, an ever-available material anchor that can be pointed to with minimal, if any, viewpoint gymnastics. The material anchor account seems to better account for example (9), especially in light of the fact that the speaker is simultaneously attributing features to *two* bodies, the mother and the daughter. Anchoring reference to other bodies by pointing to one's own body may count as a gestural practice, a particular multimodal activity type like a re-enactment or a tracing gesture (see Sidnell, 2006; Enfield, 2009, pg. 19). The notion of the body as material anchor will prove central to an account of the conceptual processes underlying the next type of body-directed gesture, body-anchors.

3.5. Body-anchors

The final type of body-directed gestures, *body-anchors*, is exemplified by cases in which the speaker indexes abstract experiential notions— whether mental, emotional, or physiological— to a certain part or region of the body. In the TSC, notions that are anchored to the body in gesture include *instinct*, *dreams*, *feeling*, and *courage*, among others (n= 146). Such notions are canonically private, unseen, and abstract, without spatial location or extent or concrete form. Yet through body-directed gestures speakers point to them all the same. *Anchoring*, as the term is used here, refers to just this spontaneous spatialization and resulting concretization.

Figure 3.6: Three frames from Example 10. Tavis produces a self-point with "I" (a), followed by a body-anchor with "comfortable" (c). Note the change in handshape between the two gestures.

In the first body-anchor example, Tavis is talking with the actress Julia Louis-Dreyfus about the types of roles she usually plays, and asks about how she came to play them:

Example 10— JL 10:21

TS: 1 how did you know-
2 when did you know
3 okay **I can- I** can do this and I'm **comfortable** playing these kinds of roles

In line 3, Tavis produces a self-point to his chest with the index finger of his right hand as he says "I can do this". As he continues to say "and I'm comfortable" his handshape changes: all his fingers become rigidly extended toward his body so that the tips make contact with his upper torso (see Figure 3.6). The gesture construes "comfortable" as a felt bodily experience, immanent in his torso. The change of handshape and the contact with the torso together perhaps express the idea of being *comfortable* as something diffuse rather than narrowly localized in the body.

In example 10 there does not appear to be any deeply meaningful relation

Figure 3.7: Three frames from example 11. The speaker produces a dynamic body-anchor gesture (b-c) in the area of his head to express the idea of getting "sucked into it".

between the *particular* indexing site— the upper-right part of Tavis's torso— of the gesture and the experiential notion— "comfort"— he anchors there. There appears to be at most a general notion at play of the torso as container for emotions and experiences. Elsewhere, the indexing site appears more critical to the meaning of the body-anchor. Different bodily regions bear culturally mediated and often highly entrenched associations. In example 11, Hill Harper is discussing a trend in which teenage boys capitalize on new technologies like text messaging to manipulate teenage girls:

Example 11— HH 07:15

HH: 1 they say things they wouldn't even have the courage
2 to say face to face
3 but a lot of these girls **get sucked into it** because they feel like they need-
4 they need to reach out

As he says "get sucked into" Harper draws both hands toward his head in slow, circular motions (see Figure 3.7). By positioning the gesture where he does in relation to his body, Harper would seem to express the idea of "getting sucked in" as a kind of cognitive struggle. The bimanual movement pattern, which is iterated but not symmetrical, seems

Figure 3.8: Three frames from example 12. Co-timed with "sick" the speaker produces a highly articulated but rare "claw" hand-shape (b-c), which embodies the emotional character of the experience she is anchoring to her torso.

to subtly enrich the meaning of the gesture, perhaps by suggesting that the struggle is drawn out or that disparate forces are involved.

Morphologies that body-anchors assume can also embody the affective character of experience. In the following example, the actor Judy Reyes asserts that, despite the inner anguish experienced by a grieving mother following the loss of her child, there is an urgent need to move on:

Example 12— JR 05:19

JR: 1 it can't be about
2 kicking and screaming kicking and screaming
3 putting herself- making herself **sick**
4 when she's forgetting about the two children that exist

Co-timed with the word "sick" in line 3, Reyes locates the mother's inner experience on her own body, specifically her upper torso, with a two-handed gesture (see Figure 3.8). The gesture characterizes the affective qualities of the experience through its rigid claw-like handshape: the fingers of both hands are flexed toward her chest in a well-articulated

morphology. But what exactly is the relation between the experience of *sick*-ness to which she refers and the shape of her hands? On a standard notion of iconicity as perceptual resemblance, the jagged claw shape could be a synaesthetic image for inner anxiety. Alternatively, the visual image of the handshape may be only incidental, the real iconicity residing in the muscular tension— a kind of kinesthetic image— the speaker feels as she forms and holds her hands in the posture. In this and other examples of body-anchors, indexical and iconic elements fuse into a seamless gestural whole.

Common to each of the three foregoing examples is the fact that the speaker anchors someone *else's* experience in his or her own body. In (10) it is the experience of the immediate addressee; in (11) it is the experience of a loosely defined segment of society, i.e "teenage girls"; in (12) it is the experience of a grieving mother. How is it that speakers can bring forth others' experiences by indicating their own bodies? As is clear from the examples, a speaker's ability to evoke another is not at all constrained by some *a priori* likeness between the speaker's body and the evoked body. Evoked bodies often differ in sex, age, and physical attributes. Apparently, the schematic similarity common to all human bodies— perhaps underpinned by an abstract body template— enables effortless leaps of perspective, from *I* to *you* to *he*[30].

[30] As informal evidence of such a conceptual body template, consider the curious fact that, at least in English, speakers are readily able to use the word "here" while indicating a location on their own body to mean at the corresponding location on *another's* body. Hanks (2005, pg. 201) briefly discusses an example of this in Yucatec Maya.

3.6. Discussion

3.6.1. Form and function of indexical gestures

Body-directed gestures offer a neatly delimited empirical arena in which to sharpen theoretical proposals about the form and function of pointing gestures. Body-directed gestures are commonplace and yet, at the same time, formally unlike the familiar prototypes of pointing. How do we account for these differences in form? Attention to variation in pointing form is not new. Kendon (2004) has proposed that the form of a pointing gesture is "systematically related to the way the object being referred to is presented in the speaker's discourse" (pg. 201). Based on naturalistic data from speakers in southern Italy and Britain, he identifies a number of morphological variants in pointing which he argues are associated with particular meaning variants. For example, he proposes that open-hand pointing is used when the object "is being referred to in virtue of its status as a symbolic, conceptual, or exemplary object" (pg. 223). Wilkins (2003) presents a similar style of analysis of pointing handshapes used by speakers of Arrernte, an aboriginal Australian language. Both these analyses posit that culture-specific, conventionalized form-meaning pairings are responsible for morphological variation in pointing. The apparent rigidity of a formal convention does not rule out the possibility, of course, that the convention is motivated in some way. Haviland (2003) has proposed, for example, that even the canonical index-finger morphology is tacitly iconic, with roots in the notion of "oneness" (see pgs. 162-3 for discussion).

Over and above the kinds of conventional pairings that Kendon and Wilkins propose, it is possible that certain general principles may be at play in determining pointing morphology. One proposal, alluded to above, is that pointing morphology is best

understood as construal. *Construal* here is meant in the same sense in which it is widely used in the cognitive linguistics literature: as evanescent meaning built on the fly, meaning which is highly dependent on local context and which arises out of the subjective cognitive processes of the language user. The morphology as construal proposal suggests partial answers to two puzzling features of self-points that distinguish them from their externally directed counterparts— namely, the fact that they are very often two-handed and the fact that they are often produced with full-hand morphologies.

Part of the answer to the puzzle may be that pointing form reflects the granularity of attention-direction intended. The canonical pointing gesture involves a particularly fine-grained type of attention-direction, one in which a well-circumscribed, focal referent (e.g. a cruller) is singled out amidst a crowded backdrop of other candidate referents (e.g. a case of donuts). The extended index finger in the canonical scenario serves to construe the region of interest as small and focal, a mere speck in the visual field. Other times *coarse*-grained attention-direction is in order, for example when indicating which of two sides of a large auditorium is meant. To use an extended index finger in such a situation would be to, as it were, misconstrue it. It is not only the absolute size of the referent that matters— it is the construal of the relation between the referent and an assumed attentional field. In producing a self-point, the speaker most often construes the self, not as a speck in the attentional field, but as a large portion of it. Morphologies that are two-handed and/ or full-handed embody this construal[31]. In addition to the granularity of attention-direction desired, other aspects of construal may motivate the use of full-handed

[31] In light of this suggestion, it is interesting to revisit Kendon's (2004) proposal that the use of an index-finger morphology when pointing to the self is associated with the self seen from "outside".

morphology. A related factor is the relationship between the *site* of the pointing gesture and its *target*. If the site and the target are roughly co-extensive— as they are, for instance, in the donut scenario— an index finger may be more apt. However, if the speaker construes the site as a subset of some larger target, a full-hand or two-handed morphology may be used. Consider again the constructed example discussed above of a person pointing to a wall that is nearby. In such a case, the site of the pointing gesture is but a tiny portion of the intended referent, the entire extended surface of the wall. Thus, if the explanation offered here is correct, we ought to expect a full-hand or two-handed pointing gesture in this case. A key virtue of this morphology-as-construal proposal is that there is no need to posit a qualitative discontinuity between self-points and externally directed, canonical points. Both are readily accounted for by appeal to general processes in how pointing gestures are realized.

In the case of self-points, then, the form used covertly construes aspects of the relation between the site, the target, and the attentional ground. In body-points and body-anchors, pointing form further— and often more overtly— reflects construed properties of the referent, such as its spatial extent, its conceptualized motion, or its affective character. Such gestures reflect a seamless fusion of iconic and indexical elements (c.f. Haviland, 2003, pg. 162 on pointing gestures as “complex semantic portmanteaux”). Indexical gestures sometimes go beyond characterizing a referent in some minimal way. Goodwin (2007) has described a phenomenon of "environmentally coupled gestures", in which gestures are layered over existing material representations, such as pointing gestures made in relation to dirt etchings at an archaeological field site (see also Hutchins [2006]). Such gestures are only intelligible, Goodwin argues, by virtue of their tight

contiguity with an indexical ground. It might be added that coupling is not limited to cases in which iconic gestures are layered over *external* material representations out there in the world. Coupling can be viewed instead as general higher-order property in which iconic properties of a gesture only make sense in light of their indexical relationship to some *demonstratum*. Very often in body-anchors, as shown above, an indexical coupling exists between an otherwise iconic gesture and the body in a way that fundamentally changes the gesture's meaning. In (11), for example, Harper's head-directed gesture when describing "getting sucked in" would take on an entirely different semantic shading if directed instead towards his stomach. In that case, it might imply an emotional or visceral rather than cognitive involvement.

Formal complexities are a hallmark of body-directed gestures, and so too are functional complexities. Recall that, according to the standard account, pointing gestures primarily serve to re-direct a listener's visual attention to a referent. Other functions are often posited or merely assumed, such as that pointing supports singular, unambiguous reference. Self-points co-produced with first-person pronouns, to take the most straightforward case, do not obviously serve either function. The audience's attention presumably does not need to be re-directed to the speaker; and, moreover, first-person pronouns are not by their nature ambiguous in the way that other deictic words sometimes are (Levelt et al. 1985). It is on the basis of such functional considerations that Pizzuto and Cabobianco (2008) have recently argued that pronominal points to the speaker and addressee— as they appear in both spoken and signed languages— ought to

be considered *separately* from other deictic gestures[32]. Specifically, they argue that since pronominal points do not induce shifts in visual attention to a target, they are best considered a different beast.

Why do people produce self-points if not to induce changes in visual attention? One important consideration is that visual attention is a coarse and imperfect proxy for attention generally. Yet the question remains why a speaker would point along with speech that is by all accounts unambiguous on its own: the word *I* should always suffice to pick out the speaker without gestural supplementation. Discourse structure offers part of the explanation. Enfield et al. (2007) have demonstrated an association between a particular feature of utterances— "location focus"— and the robustness of pointing gestures. More basically, discourse structure may condition the very presence or absence of a pointing gesture in the first place. In line with this idea, the current study observed a clear association between the very presence of self-points and contrastive focus. Self-points may serve to focus audience attention, not just on a target that is either currently unattended or that is necessary for disambiguating an utterance, but on what is currently *foregrounded* in the discourse. Further research will be required to determine whether aspects of information structure beyond contrastive focus and location focus are associated with pointing, and, moreover, whether information structure might be implicated more generally in the production of co-speech gestures of other types.

---

[32] Kendon (2010) has recently proposed a similar bifurcation, as he formulates it "between *personal pronominal pointing*, in which one points to oneself or others and *external reference pointing*, in which one points to indicate objects, locations, or directions which are external to the speaker or external to his discourse" (pg. 21). But he offers no rationale for the bifurcation.

3.6.2. Conceptual processes in body-directed gestures

Body-directed gestures involve conceptual processes of different kinds. Meriting special discussion are those processes that have not received due attention elsewhere in co-speech gesture studies, including processes of *intercorporeality*, *compression*, and *anchoring*. Such processes are manifest in cognitive activities and cognitive products of very different sorts. In the present data they are manifest as they unfold moment-to-moment in spontaneous acts of meaning-making.

Intercorporeality, as the term is used here, refers to the process of mapping fluidly from one human body to another, from *I* to *you* and back again. Bodies are an especially salient stimulus in the world; reasoning about them is thus not like reasoning about other kinds of entities in the world, such as tides and volcanoes and satellites (c.f. Wilson & Knoblich, 2005). The capacity to see correspondences between bodies— and to reason with these correspondences— is very likely basic and unreflective. Intercorporeality is evident in each of the three types of body-directed gestures described above. Self-points are occasionally produced with second and third person pronouns, such as in examples (4) and (5) above. It is crucial to note that in such examples the speaker does not appear to fully shift from one perspective to another, as happens in cases of enacting gestures (Kendon, 2004) or multimodal demonstrations (Clark & Gerrig, 1990). Instead two viewpoints are interposed: the viewpoint of the pointer and the viewpoint of the pointed-to. The speaker's body is both the speaking/ gesturing body and, at the same time, the body of another who is being referred to. Evidently intercorporeality makes possible, not just switching in quick succession between distinct viewpoints, but also the simultaneous embodiment of more than one viewpoint. Interpositions of this type

are also manifest in body-points, which are regularly produced in the course of reference to other bodies, even starkly different bodies (i.e. bodies of different sexes, ages, hairstyles, etc.) as in examples (9) and (11). Body-anchors, moreover, are very often indexed to the speaker's body when the explicit verbal reference is to generic human experience or to the felt experiences of a specific, narrated other.

Intercorporeality is not necessarily a rarefied conceptual achievement. Much evidence suggests, in fact, that it is supported by low-level, mechanistic processes. Motor resonance, a widely studied phenomenon in which viewing the motor actions of others induces activity in one's own motor system, is a particular manifestation of automatic, often subconscious intercorporeality (see Wilson & Knoblich [2005] for a review). In humans the roots of intercorporeality can be seen very early in development, for instance, in the propensity to imitate adult facial gestures (Meltzoff & Moore, 1995). In fact, intercorporeality is likely not a human-specific achievement at all. Consider that, in the fabled discovery of mirror neurons, neurons in a macaque's brain responded to the actions of a *human* experimenter, not those of another macaque. Whatever the ultimate origins of the ability, intercorporeality is manifest in a panoply of overt human communicative behaviors of varying complexity, from the character viewpoint gestures produced by young children (McNeill, 1992, Ch.11), to the pointing gestures of a severely aphasic man (Goodwin, 2006, pg. 110), to sophisticated scientific reasoning practices (Alač, 2009).

Another conceptual process endemic to body-directed gestures is *compression*, which has been described in some detail by Fauconnier & Turner (2002). In compression, what is multiple, extended in space, and protracted in time is construed as singular,

compact, and instantaneous. Humans reason best at human scale, and compression is a key mechanism by which unwieldy scenarios— scenarios that are spatially, temporally, causally, and intentionally diffuse— are brought into cognitive grasp. In support of this idea, Fauconnier and Turner discuss numerous examples of riddles, diagrams, newspaper headlines and the like in which what they call "vital relations" of space, time, identity, and causality are compressed. Spontaneous compression, unfolding on the scale of milliseconds, is likewise evident in many of the gestural examples described above. When, as in example 3, a speaker produces a point to the self while saying "we", there is a compression of speaker and relevant associates to the speaker's body[33]. When speakers produce self-points with expressions like "as a man" or "as a cancer survivor", they compress entire categories of persons to a single category exemplar. When speakers point to their own body parts in the course of reference to a *generic* human body (example [8]) or with an impersonal *you* (example [4]) they likewise compress types and tokens. Gestural compression of this sort turns a scene that is spatially, temporally, and causally diffuse into one that is human-scale and familiar, a scene in which salient discourse entities become jointly attendable in the here-now of interaction.

Compression is close cousin to a much more widely discussed phenomenon, metonymy. Here compression and metonymy are treated as *process* and *product*, respectively. Traditionally, metonymy has been defined as a rhetorical device, evident in certain linguistic constructions, in which reference is made indirectly. For instance, in the sentence "Washington pressured London into lifting trade restrictions" the names of

[33] One could argue that even the most mundane self-points co-produced with first-person pronominals reflect feats of compression— namely, a compression of personal identity and the abstract participant role of speaker at the moment of gesture.

cities— Washington and London— are metonyms that indirectly refer to the governments which the cities house. More recently, in the cognitive linguistics tradition, metonymy has been re-imagined as a general conceptual process rather than an exclusively linguistic one (Kövecses & Radden, 1998; Coulson & Oakley, 2002). In these accounts, one element of a frame serves as a "trigger" for another element of the same frame. Metonymy is thus primarily about relationships that afford cognitive access across frames or across mental spaces, and certain relationships such as part for whole, producer for product, controller for controlled arise more commonly than others. Despite the important theoretical move to acknowledge the existence of metonymy beyond language, there has been little empirical effort to study metonymies that arise in real-time, everyday behavior. Doing so invites a shift in emphasis from *metonymy*— which labels a static relation— to *compression*— perhaps the central process through which metonymies are minted[34]. An additional virtue of the shift to compression is it recasts metonymy, not as a fanciful kind of figurative language, but as a motivated outcome of the fundamental conceptual drive to "achieve human scale" (Fauconnier & Turner, pg. 346).

A final conceptual process evident in body-directed gestures is *anchoring*. Anchoring is the assignment of meaning to a particular location in space. In the minimal case, one can anchor reference to some concrete entity that is physically *outside* of the discourse context by locating it in space *within* the discourse context. Such anchoring can happen in empty space, as it does in narrative contexts when speakers make pronominal reference by gesturally assigning characters to loci (McNeill, 1992) and, even more

[34] For discussion of the relationship between metonymy and compression, see Fauconnier (2009).

pervasively in, in sign language narratives (Liddell, 2000). Other times anchoring occurs "on top of" existing material structure. In the latter case, the assignment of meaning to a certain locus may be motivated by particular features of the available material structure. Importantly, what is anchored to space is not always a discrete "entity" in the first place; it can be something altogether more amorphous, such as an idea or a category, in which case compression may also be involved. For instance, in examples (10) and (12) the experiences of being "comfortable" and "sick" are anchored to particular locations on the speaker's body. Such cases involve the simultaneous compression and spatialization of abstract ideas.

The use of the term "anchoring" above differs somewhat from how it has been used elsewhere in the cognitive semantics literature. Hutchins (2005) uses the concept of anchoring to describe a human strategy of associating conceptual and material structure, one which manifests in the pervasive use of "material anchors" (pg. 1555; see also Fauconnier & Turner [2002], especially Ch. 10 on 'Things'). Among his many examples of material anchors is the Japanese hand calculator, a device for determining which day of the week a given calendar date will fall on. In the case of the calculator, conceptual structure is projected onto the anatomy of the human hand in a way that facilitates rapid and precise computation. In Hutchins' account, the power of material anchors such as the hand calculator lies largely in the fact that they hold conceptual relations stable, making those relations ready for further conceptual elaboration.

The emphasis here is on the initial, spontaneous act of anchoring meaning in space rather than on any downstream cognitive consequences of such anchoring. Material structure is very often, though not always, involved in initial acts of anchoring, as it is in

all the examples of body-directed gestures considered here. The opportunistic use of material structure is motivated, not random. The body is a very peculiar and powerful kind of material structure, possibly unique in the disparate ways and degree to which it supports spontaneous anchoring. One reason for this power is that the body is an ever-available resource in interaction. Another reason is that the body invites compressions. Because a person belongs simultaneously to numerous categories (social, biological, etc.), the body supports compressions of diverse *types* into a single *token*. Because it is layered over with countless, culturally mediated associations between particular body parts and particular experiential notions, the body supports compressions of the abstract, ineffable, and subjective into visible, physical loci. In these respects, the body is a powerful material anchor in something more like Hutchins' sense. It constitutes a stable and highly organized field of social and experiential meanings, and a concrete, physical instantiation of the body template.

In the case of body-directed gestures, these three different conceptual processes more often act in concert than in isolation. Take example (11), in which Harper describes teenage girls getting "sucked in" by the stratagems of teenage boys. Intercorporeality enables the speaker to use his own body to model the body of someone else; compression enables disparate and probably non-specific persons— "young girls"— to be merged into a single body; compression further allows the process of "getting sucked in" to be scaled in time and space; finally, the mooring of the process of "getting sucked in" to the area around the head constitutes anchoring. In example (12), we find a similar confluence of processes. Again, intercorporeality makes possible the use of the speaker's body to stand in for that of the grieving mother. The felt experience of being "sick" is compressed and

anchored to a particular locus on her body. Thus while these processes are logically dissociable, particularly in other classes of cognitive activity, in spontaneous body-directed gestures they are frequently tangled up in a sort of conceptual bricolage. It bears noting, finally, that while processes of intercorporeality, compression, and anchoring are especially apparent in the body-directed gestures considered here, they may well turn out to be central to other pointing gestures to varying degrees.

3.6.3. Body-directed signs

Body-directed gestures have clear analogues in signed languages, forms which might be termed body-directed *signs*. In the literature on both emerging and fully established sign languages, body-directed signs— though not often grouped as such— have been shown to be a pervasive and variegated phenomenon. Kendon (1980), in his discussion of the Enga sign language of the Papua New Guinea highlands, draws a distinction between signs that are articulated in "neutral space" and those that are articulated "in relation to the body". He reports that, in his sample of 97 Enga signs, 41% were the latter type, *body signs*. He then, by way of comparison, presents an analysis of the proportion of body signs in corpora of diverse signed languages: American Sign Language (32%); British Columbia sawmill sign language (37%), and an aboriginal sign language of Queensland Australia (23%). Other studies of emerging sign languages further attest to body-directedness as a powerful and indeed pervasive semiotic property of signs (see, e.g. Kuschel, 1974).

Self-points, body-points, and body-anchors all have clear counterparts in signed languages. Where they appear, such signs invariably bear traces of the three conceptual

processes just discussed[35]. In many— if not all— sign languages, first-person pronouns are signed by pointing to the signer's torso (see Friedman [1975] on ASL). In fact, debate continues about whether self-pointing signs in sign languages are better considered part of the grammar of the language or instead examples of mere pointing, and thus analogous to co-speech gesture (see Meier, 1990; Pizzuto & Capobianco, 2008). Data presented in this study are germane to the debate, particularly as they describe the parameters along which co-speech self-points are and are not regularized. They thus provide a reference point— albeit from a single cultural-linguistic group— against which variation in self-pointing signs might be fruitfully compared. Sign languages are also replete with body-part lexicon, of course, and in ASL these partonymic signs invariably involve pointing to a part or region of the body, though not always with an extended index finger (Pyers, 2006). To my knowledge, body-anchors in sign languages have not been studied extensively in their own right, but there is little reason to doubt their ubiquity, particularly in the context of emerging sign systems. Kendon's Enga corpus, for example, includes signs for SEE— formed by pointing an index finger to the eye— and FORGET— formed by thrusting a hand outward from the side of the head— among others. Many of the noteworthy features of body-anchors in co-speech gesture described above— e.g. their interlacing of iconic and indexical elements — appear evident in these and other signs in Kendon's sample.

In the face of the seeming transparency of many of the body signs Kendon reports, it must be cautioned that they— and their counterparts in co-speech gesture— do

[35] Of course, a number of other powerful forces are involved in sign formation, for instance the need for new signs to contrast with other existing signs, a problem not faced acutely by speakers in spontaneous gesture.

not necessarily reflect natural, universal associations. They are all the more interesting in light of this fact. An especially vivid illustration of this truth is recounted by Malinowski in *The Argonauts of the Western Pacific* in describing an exchange with a Trobriand magician: "I paid him well for the formulae he gave me, and inquired at the end of our session if he had any more magic to produce. With pride he struck his belly several times and answered: 'Plenty more lies there'!" (qtd. in Senft, 1998, pg. 74). As the man's gesture makes clear, according to the Trobriand conceptualization, knowledge of magic is contained in the abdomen[36]. More recently, as part of a much larger project on the semantics of mental processes in Australian languages, Evans and Wilkins (2000) consulted Kendon's corpus of Central Australian mourning signs. They report a strong tendency for signs articulated in relation to the ear to have cognition and intellection meanings, while signs articulated near the eye tend to have perception and emotion readings. Both body-point and body-anchor signs thus potentially offer a window on how the body— and its associated experiential processes— is variably construed cross-culturally. There have been several recent linguistic studies of the construal of the body (e.g. Gaby, 2008), and it seems these might be enriched by also considering co-speech body-directed gestures. Indeed, body-directed gestures constitute a new empirical arena in which to explore— cross-cultural as well as cross-modal— variation in associations between parts of the body and experiential notions.

[36] I thank John B. Haviland for pointing me to this passage.

3.7. Conclusion

Body-directed gestures are as commonplace as they are diverse. The body is a number of things at once: it is a social body, an interactive body, an experiential body, and, of course, a biological body with parts and functions (Mol, 2002). The body, as such, is enlisted in everything from the multimodal articulation of the grammatical category of person— in both speech and sign— to the articulation of our most abstract experiential notions. Every speaker has a body, and every speaker has occasion to point to it. Together the body's multiplicity and basic-ness make body-directed gestures an illuminating case study, a fruitful microcosm in which to take up questions about gestural indexicality, its relation to speech, and its conceptual underpinnings.

References

Alač, M. (2009). Moving Android: On Social Robots and Body-in-Interaction. *Social Studies of Science*, *39*(4), 491-528. doi: 10.1177/0306312709103476.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*, 415-419.

Calbris, G. (1990). *The Semiotics of French Gestures*. Bloomington: Indiana University Press.

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

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 243-268). Mahwah, NJ: Lawrence Erlbaum.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 245-258. doi: 10.1016/S0022-5371(83)90189-5.

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, *66*(4), 764-805.

Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 25-49). Basingstoke, England: Palgrave Macmillan.

Coulson, S., & Oakley, T. (2003). Metonymy and Conceptual Blending. In K.-U. Panther & L. L. Thornburg (Eds.), *Metonymy and Pragmatic Inferencing* (pp. 51-79). Amsterdam: John Benjamins.

Dancygier, B. (2008). Personal pronouns, blending, and narrative viewpoint. In A. Tyler, K. Yiyoung, & M. Takada (Eds.), *Language in the Context of Use: Discourse and Cognitive Approaches to Language* (pp. 167-182). New York: Mouton de Gruyter.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.

Enfield, N., Kita, S., & de Ruiter, J. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722-1741. doi: 10.1016/j.pragma.2007.03.001.

Evans, N., & Wilkins, D. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, *76*(3), 546-592.

Fauconnier, G (2009). How compression gives rise to metaphor and metonymy. Plenary address given at the 9th conference on Conceptual Structure, Discourse, and Language, Cleveland, OH.

Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Floyd, S. (2010). *Discourse forms and social categorization in Cha'palaa*. Doctoral Dissertation, University of Texas– Austin.

Friedman, L. A. (1975). Space, Time, and Person Reference in American Sign Language. *Language*, *51*(4), 940-961.

Gaby, A. (2008). Gut feelings: Locating intellect, emotion and lifeforce in the Thaayorre body. In F. Sharfian, R. Dirven, & N. Yu (Eds.), *Culture, body, and language: conceptualizations of internal body organs across cultures and languages* (pp. 27-44). New York: Mouton de Gruyter.

Goodwin, C. (2006). Human Sociality as Mutual Orientation in a Rich Interactive Environment: Multimodal Utterances and Pointing in Aphasia. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 97-125). New York: Berg.

Goodwin, C. (2007). Environmentally coupled gestures. In S. Duncan, J. Cassell, & E. Levy (Eds.), *Gesture and the dynamic dimensions of language* (pp. 195-212). Philadelphia: John Benjamins.

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

Hanks, W. F. (2005). Explorations in the Deictic Field. *Current Anthropology*, *46*(2), 191-220. doi: 10.1086/427120.

Haviland, J. B. (2000). Pointing, gesture spaces, and mental maps. In D. McNeill (Ed.), *Language and gesture* (pp. 13-46). Cambridge: Cambridge University Press.

Haviland, J. B. (2003). How to point in Zinacantán. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 139-170). Mahwah, NJ: Lawrence Erlbaum.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495-514. doi: 10.3758/PBR.15.3.495.

Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of Pragmatics*, *37*(10), 1555-1577. doi: 10.1016/j.pragma.2004.06.008.

Hutchins, E. (2006). The Distributed Cognition Perspective on Human Interaction. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 375-398). New York: Berg.

Jorio, A. de. (2000 [1832]). *Gesture in Naples and Gesture in Classical Antiquity*. (A. Kendon, Ed.). Bloomington: Indiana University Press.

Kendon, A. (1980). A description of a deaf-mute sign language from the Enga Province of Papua New Guinea with some comparative discussion. Part II: The semiotic functioning of Enga signs. *Semiotica*, *32*(1/2), 81-117.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kendon, A. (2010). Pointing and the problem of "gesture": Some reflections. *Rivisti di Psicolinguistica Applicata*, *10*(3), 19-30.

Kockelman, P. (2005). The semiotic stance. *Semiotica*, *157*(1/4), 233-304. doi: 10.1515/semi.2005.2005.157.1-4.233.

Kövecses, Z., & Radden, G. (1998). Metonymy: Developing a cognitive linguistic view. *Cognitive Linguistics*, *9*(1), 37-77.

Kuschel, R. (1974). *A Lexicon of Signs from a Polynesian Outlier Island: A description of 217 Signs as Developed and Used by Kagobai, the Only Deaf-Mute of Rennell Island*. Kobenhavns: Akademisk Forlag.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Levelt, W. J. M., Richardson, G., & Heij, L. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*, 133-164.

Levy, E. T., & Fowler, C. A. (2000). The role of gestures and other graded language forms in the grounding of reference in perception. In D. McNeill (Ed.), *Language and Gesture* (pp. 215-234). Cambridge: Cambridge University Press.

Liddell, S. (2000). Blended spaces and deixis in sign language discourse. In D. McNeill (Ed.), *Language and Gesture* (pp. 331-357). New York: Cambridge University Press.

McClave, E. (2000). When "You" Means "I": Manual and Nonmanual Gestures and Shifting Participation Frameworks. In J. K. Peyton, P. Griffin, W. Wolfram, & R. Fasold (Eds.), *Language in Action: New Studies of Language in Society* (pp. 247-259). Cresskill, NJ: Hampton Press.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: Chicago University Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: Chicago University Press.

Meier, R. P. (1990). Person Deixis in American Sign Language. In S. D. Fischer & P. Stiple (Eds.), *Theoretical Issues in Sign Language Research* (pp. 175-190). Chicago: University of Chicago Press.

Morford, J. P., & Goldin-Meadow, S. (1997). From here and now to there and then: the development of displaced reference in homesign and English. *Child Development*, *68*(3), 420-35.

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Durham, NC: Duke University Press.

Parrill, F. (2009). Dual viewpoint gestures. *Gesture*, *9*(3), 271-289. doi: 10.1075/gest.9.3.01par.

Peirce, C. S. (1940). *Philosophical Writings of Peirce*. (J. Buchler, Ed.). New York: Dover.

Pizzuto, E. A., & Capobianco, M. (2008). Is pointing "just" pointing? Unraveling the complexity of indexes in spoken and signed discourse. *Gesture*, *8*(1), 82-103. doi: 10.1075/gest.8.1.07piz.

Pyers, J. E. (2006). Indicating the body: Expression of body part terminology in American Sign Language. *Language Sciences*, *28*(2-3), 280-303. doi: 10.1016/j.langsci.2005.11.010.

Senft, G. (1998). Body and Mind in the Trobriand Islands. *Ethos*, *26*(1), 73-104.

Sidnell, J. (2006). Coordinating gesture, talk, and gaze in reenactments. *Research on Language and Social Interaction*, *39*(4), 377-409.

Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*. Philadelphia: John Benjamins.

Talmy, L. (2000). *Toward a Cognitive Semantics (2 vols.)*. Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705-22. doi: 10.1111/j.1467-8624.2007.01025.x.

Whitley, M. S. (1978). Person and Number in the Use of We, You, and They. *American Speech*, *53*(1), 18-39.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*(3), 460-73. doi: 10.1037/0033-2909.131.3.460.

Xitco, M. J., Gory, J. D., & Kuczaj, S. A. (2004). Dolphin pointing is linked to the attentional behavior of a receiver. *Animal cognition*, *7*(4), 231-8. doi: 10.1007/s10071-004-0217-z.

Yoneyama Akihiko (1997) (ed.) Nihongo-syuwa jiten [Japanese Language-Sign Language Dictionary]. Tokyo: Zennihon roarenmei shuppankyoku.

Zwets, M. (2009). *The point at issue: personal pronouns*. Unpublished Masters Thesis, Department of Linguistics, Radboud University Nijmegen.

Chapter 4. Nose-pointing in Yupno

4.1. Introduction

This chapter describes a distinctive facial gesture used by the Yupno, an indigenous group of Papua New Guinea. The gesture is used most commonly in the course of pointing, but is also produced on its own as a co-speech facial gesture. Producing the form primarily involves contracting the *levator labii* muscles, in concert with muscles of the brow, such that visible creases form around the nose and lower forehead. In its deictic uses, these facial movements are finely coordinated with movements of the head. To our knowledge, use of this gestural form— which will be labeled *nose-pointing* though, again, pointing is not always involved— in Papua New Guinea has not been previously described in any detail. Kendon (1980) makes passing mention of what may well be a similar form used by the Enga, a cultural group from the nearby highlands of Papua New Guinea, but provides no specific examples or observations about its contexts of use[37]. Nose-pointing, we argue, is not merely one pointing tool among others in the Yupno repertoire. Rather, we present a provisional analysis of the Yupno nose-pointing gesture as marked by a very specific semantic shading. What distinguishes it from other forms of pointing, both manual and non-manual, is the theme of *deictic exactness* running through its different uses. In support of this idea, we consider examples of the nose-pointing form in which nothing is apparently being pointed to at all. Such *non*-deictic uses of the face are strongly associated with

[37] Kendon's (1980) mention, in its entirety, reads: "The Enga make use of the nose as an instrument of pointing, which is done by orienting the face in the direction of the point, tossing the head back slightly and at the same time lifting the nose by contracting the nasiolabial muscles" (pg. 106)

referring expressions that foreground the semantic feature of "smallness". Taken together these observations hint at a special relationship between the nose-pointing gesture and a particular conceptual theme, diminutiveness.

In the Western world, pointing is strongly identified with a specific prototypical form: the combination of extended arm and extended index finger. However, a handful of recent studies have shown that, in many cultures, index finger pointing exists alongside other kinds of pointing. Deictic use of the head has been investigated in several industrialized cultures, including English speakers in the United States, and has been characterized as a probable human universal (McClave et al., 2007). Despite such claims of universality, there is little question that head-pointing in the West is properly considered marginal; its use appears driven by opportunism rather than convention. Culture-specific conventions governing variants of manual pointing have also been described in a handful of cultures. Wilkins (2003), for example, discusses a set of constrasting conventional handshapes used by Arrernte speakers, an aboriginal Australian group, for particular communicative situations (e.g. a flat hand point for indicating cardinal directions). Wilkins further argues that among the Arrernte the index-finger extended handshape lacks the privileged status it appears to enjoy elsewhere; it is best characterized as but one of several "allomorphs" in a broader class of "one-finger pointing".

By far the best-described form of non-manual pointing gesture is lip-pointing. Lip-pointing involves protruding one or both lips, often in concert with a raising or "tossing" of the head. Examples of lip-pointing have now been described in considerable detail by Sherzer (1972), as used by the San Blas Cuna of Panama, and more recently by

Enfield (2001), as used by speakers of Lao. In fact, lip-pointing is now known to be extremely widespread geographically, with numerous— though, unfortunately, largely anecdotal— reports of its use in South America, Papua New Guinea, aboriginal Australia, Africa, and the Caribbean, among other places (for discussion of the distribution of lip-pointing, see Wilkins, 2003, pg. 174-178). Other forms of non-manual pointing have received scant attention in the literature, and the specific morphology under consideration in the present paper has not been previously investigated in any detail.

The Yupno commonly point with the index-finger— along with other handshapes, such as the full hand— but are particularly notable in their abundant use of *non*-manual pointing. Impressionistically, Yupno speakers produce deictic head movements in considerably higher frequency than do speakers of English (see section 4.7 below). Only a subset of Yupno head-points *also* involve co-production of the nose-pointing face, prompting the question of what additional pragmatic or semantic ingredient— if any— the facial action adds. Nose-pointing thus exists in Yupno communicative practice as part of diverse pointing repertoire that includes different forms of hand pointing and head pointing, and it is perhaps best viewed as an occasional "inflection" on head-pointing. In what follows, we briefly describe the Yupno cultural context before moving to discussion of details of the nose-pointing form, examples of its use, discussion of a proposed meaning and iconic basis, and a comparison of nose-pointing to lip-pointing. We conclude with discussion of the broader interest of nose-pointing, suggesting, first, that it presents an illuminating study in the relationship between gestural and linguistic meaning and, second, that it contributes to a promising line of investigation into spontaneous, co-speech facial action.

4.2. The Yupno

The Yupno occupy more than twenty villages scattered throughout the Upper and Lower Yupno Valleys of Papua New Guinea's Finisterre Range, near the border between the provinces of Madang and Morobe. At present, no roads reach the valley from the coastal urban centers; the only approach to the Yupno territory is by multiday walk from the coast, or else by single engine aircraft to the regional airstrip at the village of Teptep. Due in large part to the remoteness and ruggedness of the Yupno territory, the group's exposure to Western cultural practices has been irregular. Only a very limited influx of media, consumer products, and emissaries reach the valley from the industrial world. Tourism is essentially non-existent. Foreign missionaries and anthropologists, however, have sporadically visited the territory on-and-off since the middle of the 20th century, and it is not uncommon for Yupno adults to have traveled to Madang or Lae, the two closest coastal cities.

Yupno is a non-Austronesian language that has been only patchily described, most notably in a series of anthropological publications by Jürg Wassmann and Verena Keck. More thoroughgoing linguistic description is underway, but is incomplete (James Slotta, *personal communication*, September 8, 2010). The Yupno language is still learned as a first language throughout the Yupno territory. Children enrolled in local schools receive instruction in a combination of Yupno, Tok Pisin (the largely English-based creole of Papua New Guinea), and, to a much lesser extent, English. More complete ethnographic details about the Yupno can be found in Wassmann (1998) and Keck (2005).

4.3. Data and analysis

Data for the present study consist of high-definition video recordings of ethnographic interviews— both informal and structured— conducted in late August and early September of 2009[38]. Interview participants were unaware of our interest in co-speech gesture, and thus also of our particular interest in the nose-pointing form. In fact, through most of data collection, we were only dimly aware of the form, and were certainly unaware of just how pervasive it was. Interviews were carried out almost exclusively in the village of Gua, with some additional material obtained during day trips to the neighboring villages of Uskokop and Gangalut. The total video corpus contains more than 15 hours of conversation on a wide range of topics, from which we identified over 80 examples— produced by 13 different speakers— of the nose-pointing form for further analysis. In its deictic uses, the form is sometimes produced without speech; other times in the corpus it is co-produced with Yupno, Tok Pisin, or English. Each example of nose-pointing was analyzed for its context of use and for the fine-grained morphological details of its production, including its degree of intensity, its temporal structure, and its co-ordination with other gestural articulators.

[38] The use of high-definition video proved critical in that it allowed for a much more detailed characterization of the facial gesture than would have been possible otherwise. It also allowed for the analysis of less robust cases of nose-pointing that may have gone unnoticed.

4.4. The nose-pointing form

4.4.1. Basic facial morphology and dynamics

The nose-pointing form is produced by a contraction of the *levator labii superioris* muscles located bilaterally on either side of the nose, which raise the upper lip and slightly broaden the wings of the nose[39]. Contracting these muscles brings about a constellation of changes in the human face (see Ekman et al., 2002 for fuller description, especially the description of Action Unit 9). Most saliently, it results in horizontal folds appearing along the root of the nose and lower forehead, as well as in a slight deepening of the furrows that extend from the base of the nose downward to corners of the mouth (sometimes called the nasiolabial furrow). Brow muscles are invariably involved as well, in particular the *procerus* muscle, which both knits together and lowers the brow toward the nose, further intensifying the mesh of wrinkles. In fact, it is extremely difficult to contract the *levator labii* without also contracting the *procerus* (it is, however, possible to contract the *procerus* without contracting the *levator labii*). Finally, the form often results in a narrowing of the eye aperture and a slight opening of the mouth. Informally, the combined effect of pulling the nose upward and pulling the brow downward and inward may be characterized as an effortful *scrunching*, or *pinching together* of the face. In English similar actions are sometimes labeled as "nose-wrinkling". Of special interest is fact that this precise constellation of actions is commonly associated with the expression of disgust in the Western world (see, for example, Chapman et al., 2009), a fact which we

[39] In most every case, the action is executed symmetrically— that is, with muscles contracting on both sides of the face. Note, however, that it is anatomically possible to independently contract the *levator labii* muscles. Unfortunately, the very few examples in the corpus of what looks like possible unilateral contraction are also cases in which half the face is relatively obscured.

return to in discussion[40]. When produced asymmetrically, the action may also associated with the "snarl" (Darwin, 1998 [orig. 1872]). As it used by the Yupno, there is no evidence to suggest that the form entails the slightest negative affect.

Participation of additional muscle groups varies somewhat idiosyncratically from speaker to speaker. One clear idiosyncratic variable in the present data is that the eyes are sometimes squeezed shut and other times left open. Eye-shutting appears to vary across speakers and less so across different contexts of use. A final noteworthy feature of nose-pointing is that the production of the form results in a slightly different constellation of facial changes depending on the physiognomy of the speaker. In one speaker a robust nose point results in salient diagonal lines running down the flanks of the nose; in another speaker, the form produces deep horizontal creases under the eyes.

4.4.2. Robustness

A hallmark of pointing gestures is that they can be produced with differing degrees of intensity (Enfield et al., 2007). Bodily effort evident in pointing may be seen as a proxy for communicative effort— that is, for the intensity with which the speaker intends to reorient audience attention. Interestingly, such an intensity gradient is also considered a hallmark of facial actions (Ekman et al., 2002). Intensity is manifest in nose-pointing as gradation in at least two key parameters. The first is the degree of contraction of the muscles. Forceful contraction results in more— and more *conspicuous*— changes

[40] See, for instance, Ekman's (1980) photographic essay about his research trips to Papua New Guinea. He includes an image that is described as a spontaneous exhibition of disgust produced by a member of the South Fore, of the Eastern Highlands province (plate 16 in the book). The action is indistinguishable from intense versions of the nose-pointing form described here.

to the face. For example, there is an apparent relationship between the forcefulness of the contraction and the depth and quantity of the wrinkles that emerge. The second intensity parameter is the length of time the form is held in its *apical*, or most intense, form. Some nose-points manifest as no more than a fleeting flash over the face; others are held in place for more than a second.

4.4.3. Coordination with other articulators

Nose-pointing is very often coordinated with movements of the head. Indeed, part of what makes it evident that the form functions deictically in the first place is that the speaker's head is *re*-oriented during the preparation phase of the gesture to establish a line of sight in a particular direction. Conversely, what leads us to identify other uses of the form as fundamentally *non*-deictic is the fact that there is no concurrent re-orientation of the head and gaze. Almost all deictic cases involve a break of mutual gaze with the audience in order to establish a line of sight by rotating the head in the desired direction (but see example 4 below for an exception to this tendency). In most cases, there is then also a subtle movement of the height of the head— that is, either a lifting or lowering of the chin. This component of the movement is most often not a ballistic "toss" of the head— as has sometimes been described for lip-pointing and as Kendon characterizes the Enga nose-pointing form — but a smooth repositioning. Such repositioning serves to subtly reinforce the sense that the speaker is re-orienting to something. In a mere handful of examples in the corpus, there is also an effortful forward head thrust co-produced with the nose-pointing face. For kinematic reasons, head-thrusting seems to be easiest to execute straight forward, and, indeed, all examples involving head-thrusting involve

pointing to a target more or less directly in front of the speaker. Lastly, movements of the head also signal the offset of the nose-point: tension in the head and neck is relaxed, the form vanishes from the face, and the eyes and head re-orient to the listener or elsewhere.

4.5. Examples of use

4.5.1. Examples of form used deictically

The nose-pointing facial gesture is most commonly used deictically. That is, the facial gesture is produced along with other actions which together signal the speaker's intent to reorient the audience to a region of space. Whether the deictic use of nose-pointing is in fact considered to be the form's most protoypical use by speakers of Yupno is unclear. We consider it basic here only because of the relative preponderance of deictic over non-deictic uses in our corpus.

4.5.1.1. Example 1— Multiple available pointing strategies

The first example illustrates the use of nose-pointing in a basic pointing context: in response to a question about location (c.f. Enfield, 2001). A group of six adult Yupno men are sitting outdoors during an ethnographic interview in the village of Uskokop. Four of the men are residents of the village; the other two men are research assistants from Gua who have joined the authors for a day of interviews in neighboring villages. During a brief break in the interview, one of the assistants directs a question to the local men about the location of something in the village. Three of the local men respond immediately by pointing. A first responds by lowering his head slightly, fixing his gaze, and producing a low-intensity nose-point. He subsequently builds on this gesture by

producing a sideways head-toss, which apparently serves as an iconic supplement to the initial point. During the gesture his arms remain crossed over his knees, and though he is clearly speaking concurrently his words are unintelligible. At the same time as the first man's response, a second local man sitting to his right produces a manual pointing gesture, his left arm and index finger extended fully in the same direction. Finally, a third man also responds only a fraction of a second after the first two. He re-orients his head away from the question-asker, raises it slightly, and produces the nose-pointing form. He is holding an infant in his lap at the time of the gesture, and produces the nose-point without spoken accompaniment. The example compactly illustrates a range of pointing strategies available to Yupno speakers. In response to the same question, both index-finger and nose-pointing are acceptable. The example also illustrates contrasting head movements in coordination with nose pointing: one man coordinates the facial action with a subtle *downward* movement of the head, the other with an *upward* movement. Finally, the example shows that nose-pointing can be produced with or without accompanying speech.

4.5.1.2. Example 2— Nose-pointing and hand-pointing in sequence

The second example comes from an informal conversation with two adult Gua men in which we discuss aspects of their personal biographies. Both men are proficient in Tok Pisin, as well as English to a much lesser degree. In the example, one of the men is explaining where in the village his son was born, in a cluster of houses just outside the main settlement area of Gua. He begins his answer by pointing with his left hand, index-finger extended, explaining in English that the house was "down there" (see Fig. 4.1a).

One of the interviewers comments that the location of the house means that it must have made for quite a far walk to the local health center, which lies at some walking distance from Gua in the opposite direction. The man immediately corrects an apparent misunderstanding of the distance of the house in question, saying "Ah, not very far… just…" As he says "just" he breaks gaze with the questioner and points again to the location with his left hand (Fig. 4.1b). As his left arm reaches full extension, he leans forward, tilts his head slightly downward, and produces the nose-pointing facial gesture. He relaxes the face and returns gaze to the question-asker, but follows-up by producing a few more index-finger extended pointing strokes in the same direction. The interviewer asks for clarification. The man again breaks gaze and fixes the nose-point in place, resumes the index-finger point to the same location, and says "down there" (Fig. 4.1c). Discussion follows about whether the researchers have visited the cluster of houses in question, with the two Yupno men insisting they have. One researcher asks whether the cluster is considered part of Gua and both men nod in assent. The same speaker from before says "Just… below there." Co-timed with "just"— which is protracted and spoken with markedly high pitch— he produces yet a third iteration of the combined nose- and hand-point, almost identical morphologically to the first two: the arm extension, body lean, and nose-face are tightly coordinated to reach apex at the same time (Fig. 4.1d). He retracts the gesture and returns gaze while finishing with "below there." As the first man begins his retraction, the second Yupno man, previously quiet, offers a nose-point, thrusting his head forward and downward slightly while saying "down there".

Figure 4.1: Four pointing gestures discussed in example 2. The gestures, depicted in a-d, form an ordered sequence, each separated by a brief period of relaxation. (a) An initial manual point to a house beyond view, produced without accompanying facial action. (b) A second iteration of the manual point now accompanied by the nose-pointing face and co-produced with "just". (c) A third iteration co-produced with "down there". (d) A fourth and final iteration co-produced with "just… below there".

The example demonstrates clearly that there is no prohibition on producing manual and facial gestures together. And this fact perhaps provides a clue that the hand-point and nose-point may do different kinds of interactive work. Especially interesting to note is that the speaker initially locates the house with an index-finger point only. It is only after a failure to recognize the first bid that the man supplements the hand gesture

with the nose-point. Very few of the examples in the corpus involve such a series of mixed-articulator points to the same referent. However, what examples we do have— including the present one— do *not* suggest that hand-pointing is considered an "upgrade" from nose-pointing, i.e. that nose-pointing is a reduced or less effortful form of pointing, which can thus be augmented with an index finger point if need be (see Enfield, 2001, pg. 198-202, for discussion of sequences of hand-pointing and lip-pointing). Nor does hand-pointing appear to constitute a downgrade from nose-pointing, however. It seems, instead, that the two may make different contributions to the utterance.

4.5.1.3. Example 3— Variation in form across speakers

In the next example, the researchers are having an informal conversation with an anthropologist and a group of four Yupno men about geography and culture in the village of Gua. The anthropologist is seeking a translation of the Tok Pisin word *daunbilo* (meaning 'downwards') into Yupno. A first man responds by reorienting his head in the direction of the macro-scale drainage of the valley toward the sea, slightly lifting his chin, and saying "omoden" (a Yupno spatial adverb meaning roughly 'downhill') while producing the nose-pointing form. A second, older man responds immediately afterward (though he does not appear to have seen the first man's gesture) by also supplying the Yupno word "omoden". He also coordinates a slight chin lift with the gaze re-orientation and produces the nose-pointing form. In contrast to the first, the second speaker's version of the face involves noticeable eye-shutting. One of the interviewers asks for a clarification of the Yupno word, and immediately both men, one after the other, reproduce the word "omoden" along with nose-pointing gestures. As in the first iterations

of the gesture, the second man's form involves conspicuous eye-shutting while the first man's form does not.

4.5.1.4. Example 4— Nose-pointing to the abstract

Example 4 comes from the same conversation as Example 2, a conversation about the birth of a Yupno man's son. One of the interviewers asks somewhat playfully whether his son's birth was "kalip si ngan" (a Yupno expression meaning 'in the distant past'). The answer is very clearly no, as the boy is known by all participants in the interaction to be no more than three years old. The man responds by rejecting the characterization with a smile, saying "It's not…". He trails off and then, after a brief pause, continues in a mixture of English and Yupno: "It's *apmasoq*. Or, it's some times— some years only going back. I can remember it." Tightly co-timed with the production of the Yupno word *apmasoq* the man produces the facial gesture while pointing with his head downward and slightly to his left (see Fig. 4.2a). Curiously, this appears to be the lone example in the corpus in which a speaker maintains gaze to the interlocutor while nose-pointing elsewhere. A possible explanation is the metalinguistic nature of the utterance in question: the speaker is nominating a particular word well monitoring closely how it is received. The Yupno word *apmasoq* is formed by joining the word for yesterday (*apma*) with the diminutive (*-soq*) and in this context may be loosely translated as "just yesterday". The example shows quite clearly that nose-pointing can be used to point out the abstract— such as, in this case, temporal entities— as well as concrete entities, places, and directions.

Figure 4.2: Two examples of the nose-pointing face produced by the same speaker at different points during an interview. (a) The speaker produces the face in the course of head-pointing to his left while saying "apmasoq" (yesterday + DIMINUTIVE) (example 4 in text). (b) The speaker produces the face without pointing. The face is narrowly co-timed with "just" and occurs in the context of an apparent pragmatic hedge (example 8 in text).

The speaker also appears to reproduce the nose-pointing form— this time without reorienting the head— soon after the initial production. After saying "apmasoq" his face relaxes and remains relaxed through a brief pause. But he then reproduces the face in less intense form as he says in English "some", maintaining gaze on the interviewer. We now move to consider possible explanations for such apparently *non*-deictic uses of the nose-pointing face.

4.5.2. Examples of form used non-deictically

Yupno speakers produce the same— or at least *superficially* identical— form at times when not pointing to anything. In the above examples, the fact that the speaker is producing the face as part of a pointing action is evident in the effortful *re-orientation* of

the head and gaze to some part of the environment. But in the examples considered next, no such reorientation is in evidence, and the accompanying speech further reinforces the reading that no pointing is in fact intended. With very rare exceptions, examples of the facial action without pointing seem to be co-produced with speech[41].

4.5.2.1. Example 5— with Yupno word meaning *short*

In this example, the facial form is produced three times in short succession, each time with a different degree of intensity and each time with the same word. A young man is attempting to explain the meaning of the Yupno phrase *sigak pasispmasoq* (a phrase meaning 'a very short season'). Unlike the examples discussed above, his explanation occurs as part of a structured interview in which Yupno speakers were asked to explain the meaning of various time-related Yupno expressions (results of this study are discussed in Núñez, Cooperrider, Doan, and Wassmann [*submitted*]). The man repeats the phrase several times to himself before launching an explanation. He produces the face a first time tightly co-timed with the word "pasipmasoq", with notable eye-shutting and no evident head reorientation. When the interviewer repeats the word for confirmation, the man repeats the first part of form— "pasipma"— this time more slowly and with greater intensity in the facial action. He concludes the sequence by repeating "pasipmasoq" twice more, adding a subtle version of the facial gesture (but now without

[41] In only two instances in the corpus, the face is produced without apparent pointing *and* without accompanying speech. (Recall that deictic uses of the gesture are quite commonly produced without speech.) Both instances come from an outdoor conversation in an open marketplace during which two different Yupno adults— at separate points in the discussion— can be seen, it seems, to address the nose-pointing gesture to a small child who is off camera. Such uses cases are of considerable interest, but for now we withhold any interpretation.

concurrent eye-shutting) to the final version. The Yupno word *pasipmasoq* is composed of the word *pasipma* (meaning 'short') + *-soq* (diminutive) and can be glossed as 'very short'. It is used of objects with concrete spatial extent, as well as of objects with temporal extent such as, in the present case, a season. Words indicating relative size appear to be strongly associated with the nose-pointing face, as the examples below further demonstrate.

4.5.2.2. Example 6— with Tok Pisin word meaning *small*

The next example occurs as part of an informal conversation with two Yupno men about the village of Gua and its history. During the conversation, conducted in Tok Pisin, one of the men relates a story from his childhood. He describes how everyone in Gua was happy to have the arrival of an anthropologist— everyone including "ol timbuna, liklik man, bikpela lapun man". The list can be glossed in English as "the ancestors, boys, and big old men". Each item in his three-term list is co-produced with a gesture. As he says "ol timbuna" he twists his body to his left, with his right arm outstretched (see Fig. 4.3a); as he continues with "liklik man" he now twists his body to his right and produces an intense version of the facial gesture, his eyes completely closed and his mouth slightly open (see Fig. 4.3b); as he finishes with "bikpela lapun man" his face changes rapidly— the eyes now open wide and the brow lifts— and he raises his right hand high above his head (see Fig. 4.3c). The timing of the nose-pointing face with *liklik* is precise in both onset and offset, lending weight to the inference that the facial gesture in this example is narrowly associated with that particular modifier and not with, for instance, the full list.

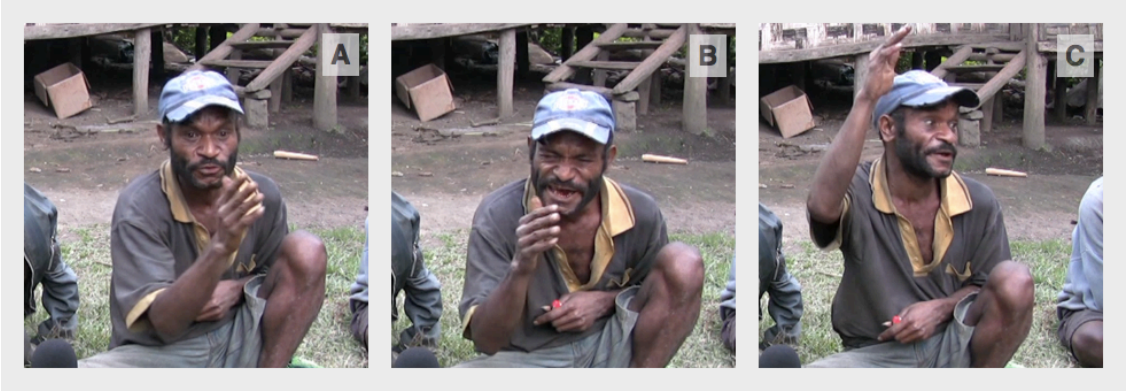

Figure 4.3: A Yupno speaker relating a story in Tok Pisin. He produces a sequence of three gestures, each linked to part of a three-term spoken list. (a) The speaker produces a manual gesture and associated head movement with "ol timbuna" ('the ancestors'). (b) The speaker twists his torso to his right and produces the nose-pointing face while saying "liklik man" ('little men'). (c) The speaker raises the torso and right hand, and expands the face and brow while saying "bikpela lapun man" ('big old men').

4.5.2.3. Example 7— with English word *little*

In the next example, a speaker provides explicit, unsolicited meta-linguistic commentary on the Yupno morpheme *–joq*, an allomorph of the diminutive form *–soq* given above. The example occurs in the context of one of the structured interviews mentioned previously, in which we are asking questions about the Yupno understanding of time. A group of men have been asked to explain the Yupno phrase *abjuk duma donjoq* meaning 'not now but a bit later'. The word *donjoq* is made up of *don* (meaning 'later') and *-joq* (diminutive). One of the men, a field assistant who has been translating parts of the conversation for us, offers the following explanation in English: "Our word *joq*. A little. Not beyond… a little bit". As he first says "a little" he produces a low intensity version of the face; co-timed with "beyond" he produces an upward head gesture, brow raised; finally, as he says "a little bit" seconds later he reproduces the facial gesture, upgrading the intensity and now coordinating it with a subtle forward head

thrust. In both cases the nose-pointing face is produced without reorienting the head and while holding gaze on the interviewer. Again, the coupling between the speech and facial action is precise in both onset and offset, reinforcing the interpretation of a narrow lexical meaning for the gesture rather than phrasal or message-level meaning.

4.5.2.4. Example 8— with English word *just*

In rarer instances in the corpus speakers produce the nose pointing face in close concert with words that do *not* explicitly convey size information. In the following example, the speaker— also the speaker in examples 2 and 4— makes the gesture while responding to a question about where he first met his wife: "When I was- when I was at Teptep doing just a labor of M-A-F— M-A-F company". Tightly co-timed with "just" the speaker produces the face along with a small amplitude headshake, all while maintaining gaze on the questioner (see Fig. 4.2b). Together the use of the headshake and the word *just* are suggestive of hedging (see Kendon [2002], pg.177-179 on headshakes as expressions of speaker uncertainty). The reason for the hedge is not entirely clear, but the speaker could be downplaying the importance of this particular stint of work in the context of his personal history. Importantly, the form itself is morphologically indistinguishable from other versions of the face produced by the same speaker in both pointing and non-pointing contexts (compare Fig. 4.2b with Fig. 4.2a).

4.6. A proposed semantic theme: The diminutive

What does the nose-pointing facial gesture mean? The fact that the form is produced as an occasional— but by no means obligatory— adjunct to head-pointing

suggests that its meaning is not reducible to or co-extensive with pointing. Rather, the addition of the facial action would seem to layer on some semantic ingredient or inflection to the base head-pointing form. The fact that the identical form is also produced *without* pointing suggests that whatever meaning it has may be profitably characterized independent of its deictic function. It is of course possible that the deictic and non-deictic uses of the facial form are gestural "homonyms" (c.f. Sherzer, 1972, pg. 118)— that is, superficially similar forms with no semantic commonality binding them. Homonymy is an unsatisfying explanation, however, for at least one important reason. Very few conventional facial expressions— that is, beyond the so-called "basic expressions of emotion"— appeared to recur in our video data[42]. It would be surprising if the two conventionalized non-affective facial gestures used by the Yupno— one for pointing and the other for denoting smallness— just happened to use exactly the same form. In the same way, it might be remarkable if a culture with an extremely limited inventory of manual emblems used exactly the same handshape to express two unrelated meanings.

Perhaps a more compelling explanation is one of gestural "polysemy"— that is, the idea that a shared semantic theme runs through the superficially disjoint contexts of use[43]. Importantly, note that such a shared semantic theme is by no means necessarily an

---

[42] It is of course completely possible that some conventional facial gestures escaped our notice, either for reasons of subtlety or rarity.

[43] Consider as an example of gestural polysemy among speakers of American English three gestures that are nearly identical in form: the 'just a minute' gesture, made by holding the index finger up for inspection; the 'nomination deictic' (Kendon, 2004, pg. 142), a discourse-related gesture that marks what is being said as of singular importance, made in the same way; and the gesture for 'one'— as in "I'll take one of those"— again

element of Yupno speakers' conscious knowledge about their own communicative practices (indeed, whether it is is an important question for further research). Given its frequent co-occurrence— and fine-grained temporal co-production— with words denoting small size, a clear candidate semantic theme for the facial gesture is the diminutive. In the above examples, co-produced smallness-related speech includes the English phrase *a little bit* (example 7); the Tok Pisin phrase *liklik man*, or boy (example 6); and the Yupno word *pasipmasoq*, comprising the word for short and the diminutive affix (example 5). Additionally, in other examples in the corpus not discussed above, the form is produced with semantically related terms such as Tok Pisin *manki* or 'boy', English *some* in the sense of few, Yupno *aminjoq*, made up of the word for man (*amin*) plus the diminutive (*-joq*), English *small*, *near,* and *little*, and at least one additional instance of co-production with the Yupno word *pasipmasoq* (produced by a different speaker from example 5). From this profusion it seems beyond doubt that the form carries diminutive meaning in certain contexts. The interesting further question is whether a diminutive semantic core might somehow also account for the many cases in which the form is used deictically, as well as for the non-deictic cases (e.g. example 8) in which smallness is not explicitly referred to.

Diminutives take on a wide range of meanings cross-linguistically. Based on a large sample of languages, Jurafsky (1996) has presented an analysis of the diminutive as a radial category. He identifies the semantic core of the category as CHILD/ SMALL and describes a number of widely attested extensions from this core, each driven by different

---

involving the same form. A notion of 'oneness' or 'singularity' undergirds these three uses, certainly, though the link may not be consciously accessible to most speakers.

mechanisms of semantic change. Among these extensions are endearment, approximation, intensity/exactness, imitation, and femaleness. Not every language that has a morphological or lexical diminutive instantiates all of these attested extensions, of course; most will instantiate only a handful. Of potential interest to the Yupno nose-pointing case are the *intensity/ exactness* extensions of the diminutive because these extensions appear to be associated with deixis. As Jurafsky describes, languages around the world use the diminutive with deictic adverbs that imply spatial extent in order to denote a narrowing of that extent. He writes: "Deictic physical location is viewed as a region in a line or a plane; diminutivization of this region converts it to a point" (pg. 550). He adds that, via the cross-linguistically pervasive metaphor of TIME is SPACE, words expressing *temporal* deictic extent, such as *now* can likewise be diminutivized. These semantic facts suggest the compelling possibility that the Yupno facial form, when layered atop head-pointing, contributes a shading of intensity/ exactness. In English an analogous function is fulfilled lexically by adding words *right* ("right here", "right now") and somewhat more restrictedly by *just* ("just then", "just above") to deictic and other spatial words (e.g. *next*, *inside, below*). In the English case, it would appear that the use of intensifying lexical material is as much a matter of pragmatics as semantics. *Right* and *just* do not necessarily encode precision in an objective sense, but rather precision against the backdrop of a presumed attentional ground[44]. If English is any guide in this respect, the Yupno facial gesture may plausibly enact a speaker's construal of a referring act as *precise* against the backdrop of a presumed attentional ground.

[44] Put another way, the supplementation of deictic words with intensifying material seems to imply, whether preemptively or correctively, that the audience's attention is focused too broadly.

Lending support to the above proposal is the fact the Yupno language uses the diminutive as a marker of deictic intensity/ exactness in precisely the way Jurafsky describes[45]. The Yupno diminutive is realized as a morphological suffix, with three allomorphs: *-soq* after a vowel, *-joq* after a bilabial/ coronal, and *-goq* after a velar. Though no detailed study has yet been conducted, the Yupno diminutive appears to be quite productive. Examples of the exactness use of the morphological diminutive include *abjukgoq* (now + DIMINUTIVE: "right now") and *odonjoq* (here + DIMINUTIVE: "right here"). Further support for the idea of the facial gesture as a marker of "exact" pointing is found in several of the above examples. In example 2, for instance, the speaker first produces a pointing gesture with no facial accompaniment and then, only upon failing to be understood, reproduces the manual pointing gesture in concert with the facial action. The fact that the nose-pointing form appears *after* the manual form alone has previously failed to succeed suggests that it functions to upgrade referential specificity (again, see Enfield, 2001, pg. 198-202). In example 4, the form is co-produced with a word involving the morphological diminutive— *apmasoq*— and is used in the sense of 'only yesterday'. The example illustrates that the precision of an indexical act is not simply an objective matter of how precisely a speaker is pointing something out; after all, in this case, what is pointed to is not visible at all. Instead, the example appears to involve a construal of the audience's attention as too diffuse when focal attention is required.

A semantic nucleus of diminutiveness thus plausibly motivates both the overtly size-related and deictic uses of the facial gesture. Curiously, an additional sense of the

[45] Linguistic facts about the diminutive presented in this section are based on field observations made by James Slotta (James Slotta, *personal communication*, October 18, 2010).

diminutive— pragmatic hedging— appears to possibly motivate other uses of the facial gesture in the corpus. In example 8, for instance, the speaker tightly co-times the facial gesture with the English word *just*, used here as a hedge marker. (Note that, interestingly, in English the word *just* functions in both the exactness sense and the hedging sense.) However, this interpretation is made cautiously; it is offered as a suggestion for further lines of inquiry into the Yupno facial gesture rather than as a well-supported finding.

4.7. Form-function fit: Nose-pointing in cultural and linguistic context

A further question that arises is that of the *fit* of the form to the particular diminutive meaning proposed. Is the form-meaning pairing a motivated match, or is it arbitrary? If motivated, what might be its iconic roots? Moreover, what cultural factors may be implicated in the selection of this particular mapping of form to function? At least two iconic bases seem plausible, and they are not mutually exclusive. First, the form could be a stylization of looking at something that is difficult to see because it is but a speck in the visual field— i.e. a form of exaggerated squinting[46]. Nose-pointing very often involves some narrowing of the eye aperture, after all. A second possibility is that the form could have roots in the kinesthetic sensation of clenching or scrunching together the face so as to make it smaller. Regardless, an important factor in the cultural "selection" of the form is no doubt its *gradability*. The nose-pointing gesture— like all facial actions (see Ekman et al., 2002)— can be produced on a gradient of intensity. At

[46] Interestingly, Wilkins (2003) mentions Arrernte speakers' use of a stylized squint along with lip-pointing "when large relative distance is being indicated, and when an object is particularly close and available" (pg. 187).

the same time, smallness is clearly a gradient feature as well[47]. The point is crucial: the use of a gradable form to mark a categorical meaning (e.g. the use of a gradable face to mark gender, for instance) or the use of a categorical form to mark a gradable meaning (e.g. the use of a categorical handshape to mark smallness) would hardly work as elegantly. It is perhaps for this reason that there appears to be some tendency for sign languages to recruit facial actions to convey size information (Emmorey, 2002, pg. 49-51; Zeshan, 2001, pg. 163). Of course, it should be cautioned that gradability is only one factor in the selection of a morphology for a particular meaning.

The above reasons do not speak definitively to the question of why the Yupno might select the *face* rather than the hands for this meaning cluster. An additional part of the answer is that the head has a different status as a gestural articulator in Yupno communicative practice. This is evident, first, in the sheer frequency with which Yupno speakers gesture with their heads compared to English speakers. Certainly, English speakers make deictic head movements (McClave, 2000), and may well do so in higher frequencies than currently appreciated. In our data, Yupno speakers are quite commonly seen to make several head gestures in sequence even while their hands are fully available. This different status is further evident in the use of the head in qualitatively different ways by Yupno speakers. It is used not merely deictically, but also iconically— for example, to trace paths of motion— as well as in ways suggestive of discourse-structuring functions. Additional analysis of video-recorded conversation in a broader range of settings will be required to validate these impressions.

---

[47] Note that, at least in some languages with morphological diminutives such as Spanish, it is possible to pile up diminutives one on top of another.

4.8. Comparison to lip-pointing

Nose-pointing differs in key respects from lip-pointing, the currently best-documented form of non-manual pointing, and these differences go beyond their superficial morphological differences. Lip-pointing is broadly distributed geographically, suggesting independent cultural invention in different parts of the world (c.f. Wilkins, 2003). To the best of our current knowledge, nose-pointing is much more limited in distribution. Kendon's earlier report of a similar form in use by the Enga raises the intriguing possibility that the gesture is used areally and is not restricted to the Yupno. But we know of no reports to suggest it is in use outside of Papua New Guinea.

One conspicuous difference between nose-pointing and lip-pointing is that, in lip-pointing, the speech articulators are obviously affected, whereas in nose-pointing they are much less so. Enfield (2001, pg. 200) makes reference to lip-pointing while speaking, but the action almost certainly changes the acoustic qualities of the co-produced speech. Nose-pointing, on the other hand, is free to operate as a truly *co-speech* gesture, as it does in nearly all examples described above: its production appears to minimally impact the accompanying speech. Another clear difference is that, assuming the reading of gestural polysemy we have proposed, the nose-pointing face can be produced independently of pointing. No examples of the lip-pointing face have been presented which do not also involve pointing, though in certain cases the deictic function is certainly less transparent (see Sherzer, 1972 for discussion).

Enfield (2001) hypothesizes that the formation of the lip-point serves to "switch on" the real deictic vector of the action: gaze[48]. Observations offered in support of this interpretation include the fact that the lips and gaze are always aligned in lip-pointing. Lip-pointing is thus cast as an elegant solution to a basic problem of face-to-face interaction, the problem of marking the difference between *incidental* gaze and *deictic* gaze. It would seem at first blush that nose-pointing may also be fruitfully conceived of as a deictic switch— as a way of "doing focused looking" in Kendon's (2009, pg. 359) terms. Such an interpretation does not fit well with our observations, however. In at least one example in the corpus (example 4), the nose-pointing face is done with clear disalignment between the head/nose and the gaze. In other cases— not discussed— the nose pointing face is produced in the course of head-pointing to referents behind the speaker. Further, given the distribution of the form across pointing and non-pointing contexts, it would be odd to propose that in pointing contexts it serves as a gaze switch while in non-pointing contexts it does not. Thus the meaning of nose-pointing we have proposed above seems, on the whole, to not fit well with the interpretation that the face functions to switch on gaze.

[48] Note, however, that in lip-pointing there is also clearly a projection of a body part in space— that is, a movement toward the intend referent, which Kendon (2004) has maintained is a criterial feature of pointing. In nose-pointing, by contrast, the actual facial movement involved is more or less orthogonal to any vector that is "projected" toward the intended referent. In other words, the nose is not actually doing the pointing— it is not "moving toward" the referent.

4.9. Conclusion

The present study has described the use of a distinctive facial action by the Yupno of Papua New Guinea's Finisterre Range. The action consists of a scrunching together of the face, which results in the formation of visible wrinkles on the nose and forehead, as well as— secondarily— a narrowing or shutting of the eyes. Dimensions of variation in the form include its intensity, how it is coordinated with other gestural articulators, and whether or not it is co-produced with speech. In the video data analyzed, the form is used most frequently as a supplement to head-pointing, but it is also commonly deployed in isolation as a conventional gesture used to characterize tiny things. Underlying these superficially disparate deictic and non-deictic uses we have proposed a semantic theme of *diminutiveness*. According to this proposal, when used as part of a pointing ensemble, the form contributes an ingredient of deictic exactness or intensity; when used apart from pointing, the form adds a gestural and thus analogue complement to verbal characterizations of size. Much remains to be investigated about the gesture and its distribution. Of particular interest is the question of just how systematic and overlapping the links are between the Yupno morphological diminutive and the observed facial action. Is the distribution of the gesture precisely co-extensive with the distribution of the linguistic marker? The Yupno use the diminutive to express endearment (J. Slotta, *personal communication*, October 18, 2010), for instance, but it is unknown whether they also use the nose-pointing face in such contexts. Also of interest are Yupno metacommunicative theories about the gesture and its meanings, such as whether it is readily recognized as polysemous and which of its uses are considered prototypical.

The interest of nose-pointing goes well beyond its place in the cabinet of cross-cultural gestural exotica. We argue that it stands to throw light on important theoretical discussions both within and beyond gesture studies. For one, it provides an illuminating case study in gestural meaning and how it intertwines with linguistic meaning. On the one hand, the gesture is clearly conventionalized: it is a particular gestural form used by a particular group for particular communicative functions. Yet it does not strictly fit the profile of an emblematic or quotable gesture, at least not according to criteria that have been laid out elsewhere (see McNeill, 1992, pgs. 56-65). At least in its non-deictic uses, it is not commonly produced without speech; it appears to admit a measure of morphological variation; it does not fall in the narrow band of speech act functions into which emblems cluster; and, finally, it does not have a narrow lexical meaning but rather a branching family of meanings. Gestural polysemy, in general, has not been widely researched. Kendon (2004) has recently developed the notion of a gesture family, a cluster of morphological cousins organized around themes such as negation and offering. It is thus tempting to speak of nose-pointing as a family of related gestures. Note, however, that Kendon's gesture families are not organized around properties of referents, but around what kinds of discourse-related or "pragmatic" functions such gestures serve.

Nose-pointing also contributes to an oddly unbalanced body of research on the role of the human face in communication. Previous work on facial movements has cleaved tightly to the framework first initiated by Darwin— and later elaborated by Ekman and colleagues— on the face as a canvas for the expression of emotions. It is striking to note, in connection to this framework, that the nose-pointing form described above is identical to the putatively universal expression of disgust (for discussion of the

facial dynamics of disgust, see Chapman et al., 2009). Disgust is very commonly included among the five or six basic emotions (Fridlund, 1994). It has even been proposed recently that recognition of the expression is supported by a dedicated brain area (Phillips et al., 1997). What is more, the form of the disgust expression itself has been argued to be not arbitrary but adaptive, with functional roots in sensory rejection (Susskind et al., 2008). We did not study the Yupno expression or recognition of disgust, but the existence of a commonplace facial action that recruits the same form raises the intriguing possibility that the Yupno might express disgust differently.

Nose-pointing is very often a *co-speech* gesture and exhibits clear links to the Yupno language. The present study thus contributes to an embryonic but promising line of investigation into co-speech facial action. Several commentators have noted the special interest of such research (e.g. Fridlund, 1994), but only a handful of studies have sought to rigorously characterize kinds of communicative facial actions that fall outside the expression of the basic emotions (e.g. Chovil, 1991/2). This empirical gap is due in part to the fact that studying facial actions poses analytic challenges; they are small in overall amplitude, they vary greatly in intensity, and they form and fade very rapidly. Much remains to be learned about the *type* and *granularity* of systematicities that may hold between facial action and speech, as well as about any cross-cultural variation in these systematicities. Hints in the literature suggest that certain cultures may operate with a specialized repertoire of co-speech facial actions (e.g. Levinson, 2010 on Yélî Dnye).

Finally, the present study enriches our understanding of pointing. Current evidence strongly suggests that humans everywhere point, but that does not mean that pointing is everywhere the same. What appears to vary are which gestural articulators are

preferred, and in what particular morphologies, as well as the assignment of certain pointing morphologies to certain communicative niches. Pointing— while certainly a basic tool in human social interaction (Kita, 2003)— appears to be subtly refashioned and elaborated anew in different cultural contexts, and there may well be dimensions to its variation that lay as yet unexplored.

I gratefully acknowledge the permission of my co-author, Rafael Núñez, to include Chapter 4, ‘Nose-pointing in Yupno’, in the present dissertation.

References

Chapman, H. A, Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: evidence for the oral origins of moral disgust. *Science*, *323*(5918), 1222-6. doi: 10.1126/science.1165565.

Chovil, N. (1991/2). Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction*, *25*, 163-194.

Darwin, C. (1998 [orig. 1872]). *The Expression of the Emotions in Man and Animals*. Oxford: Oxford University Press.

Ekman, P. (1980). *The face of man*. New York: Garland Press.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System*. Salt Lake City, UT: Research Nexus.

Emmorey, K. (2002). *Language, Cognition, and the Brain: Insights from Sign Language Research*. Mahwah, NJ: Lawrence Erlbaum.

Enfield, N. J. (2001). "Lip-pointing": A discussion of form and function with reference to data from Laos. *Gesture*, *1*(2), 185-211. doi: 10.1075/gest.1.2.06enf.

Enfield, N. J., Kita, S., & Ruiter, J. P. de. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722-1741. doi: 10.1016/j.pragma.2007.03.001.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.

Jurafsky, D. (1996). Universal Tendencies in the Semantics of the Diminutive. *Language*, *72*(3), 533-578.

Keck, V. (2005). *Social discord and bodily disorders: Healing among the Yupno of Papua New Guinea*. Durham, NC: Carolina Academic Press.

Kendon, A. (1980). A description of a deaf-mute sign language from the Enga Province of Papua New Guinea with some comparative discussion. Part II: The semiotic functioning of Enga signs. *Semiotica*, *32*(1/2), 81-117.

Kendon, A. (2002). Some uses of the head shake. *Gesture*, *2*, 147-182.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kendon, A. (2009). Language's matrix. *Gesture*, *9*(3), 355-372. doi: 10.1075/gest.9.3.05ken.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum.

Levinson, S. C. (2010). Questions and responses in Yélî Dnye, the Papuan language of Rossel Island. *Journal of Pragmatics*, *42*(10), 2741-2755. doi: 10.1016/j.pragma.2010.04.009.

Levy, E. T., & Fowler, C. A. (2000). The role of gestures and other graded language forms in the grounding of reference in perception. In D. McNeill (Ed.), *Language and Gesture* (pp. 215-234). Cambridge: Cambridge University Press.

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, *32*(7), 855-878. doi: 10.1016/S0378-2166(99)00079-X.

McClave, E., Kim, H., Tamer, R., & Mileff, M. (2007). Head movements in the context of speech in Arabic, Bulgarian, Korean, and African-American Vernacular English. *Gesture*, *7*(3), 343-390.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: Chicago University Press.

Núñez, R., Cooperrider, K., Doan, D., and Wassmann, J. (s*ubmitted).* Topographic construals of time in an indigenous group of Papua New Guinea.

Phillips, M. L., Young, a W., Senior, C., Brammer, M., Andrew, C., Calder, a J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*(6650), 495-8. doi: 10.1038/39051.

Sherzer, J. (1972). Verbal and nonverbal deixis: The pointed lip gesture among the San Blas Cuna. *Language in Society*, *2*, 117-131.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, *11*(7), 843-50. doi: 10.1038/nn.2138.

Wassmann, J. (1993). *Das Ideal des leicht gebeugten Menschen: Eine ethno-kognitive Analyse der Yupno in Papua New Guinea [The ideal of the slightly bent human: An ethno-cognitive analysis of the Yupno in Papua New guinea]*. Berlin: Dietrich Reimer.

Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 171-215). Mahwah, NJ: Laurence Erlbaum.

Zeshan, U. (2003). Indo-Pakistani Sign Language Grammar: A Typological Outline. *Sign Language Studies*, *3*(2), 157-212. doi: 10.1353/sls.2003.0005.

Chapter 5. Links between pointing and language

5.1. Links between pointing and language

Action is a ubiquitous component of utterances (Kendon, 2004). When speakers refer to things— whether seen or unseen, near or far, concrete or abstract— they produce utterances that are composites of speech and bodily action. The three empirical chapters of this dissertation have considered a particular family of actions— *pointing gestures*— and have sought to characterize several dimensions of their relationship to spoken language. Relations between speech and gesture generally have attracted much theoretical attention in recent decades, and countless empirical reports have borne out the idea that the two channels are tightly partnered in both timing and meaning. In the case of pointing gestures, however, the nature of this partnership has most often been taken for granted, the details of it left largely unexplored. This dissertation begins to address this gap in the empirical literature; it aims throw light on the texture and nuance in a behavior that is often mistaken for a monolith.

In what follows I revisit three specific issues outlined in the introduction about pointing-language relations: first, the relation between pointing gestures and deictic words; second, morphological variety in pointing and how this variety relates to speech; and, third, the lexical affiliates of pointing gestures. I consider what contributions the empirical chapters of the dissertation have made toward understanding each issue, as well as what remains poorly understood. Where appropriate I suggest particular predictions to be tested and possible courses for further research.

5.2. Reprise

5.2.1. Reprise: Pointing gestures and pointing words

The affinity between pointing gestures and verbal deixis has been remarked for more than a century. Deixis is traditionally divided into different domains, including space (e.g. *here, there*), time (e.g. *now, then*), and person (e.g. *I, you*) (Lyons, 1977; see also Levinson [2004] for additional discussion of "social deixis" and "discourse deixis"). Verbal deictics as a broad and fuzzily bordered family are considered functionally akin to pointing and, indeed, are occasionally labeled "pointing words". This functional kinship lies in the fact that both pointing gestures and deictic words exemplify indexicality, one of Peirce's three modes by which signs stand for their objects[49]. Indices stand for their objects by means of some spatio-temporal connection. Verbal deictics derive their meaning from spatio-temporally situated facts of their production, such as who is speaking, where, and when. Relative to other verbal referential devices, they are taken to be *semantically* impoverished in this way, furnishing only the most barebones characterization of their referents. Pointing gestures, likewise, derive their meaning from the spatio-temporal facts of their production. Relative to other gestural referential devices, they are taken to be *iconically* impoverished[50], providing at most a minimal and highly schematized "likeness" of their referent (see discussion of covert iconicity below). Both pointing words and pointing gestures owe their interpretability foremost to common ground (Clark, 1996), including elements of mutual knowledge, shared construals of

[49] Though traditionally taken as paragons of indexicality, both pointing gestures and pointing words are best considered, borrowing Enfield's (2009, pg. 15) terms, "symbolic indexicals" to the extent that they combine conventional and deictic elements.
[50] Recall Kendon's framing of pointing gestures as gestures "dominated by the deictic component almost to the exclusion of everything else" (2004, pg. 205).

space, and perceivable concurrent action. It is very often the case that the interpretability of pointing and verbal deixis depends, in fact, on their *co*-production.

Importantly, though both gestural pointing and verbal deixis are paradigmatically indexical in terms of their semiotic function, it does not follow that all types of *verbal* deixis are co-produced with *gestural* deixis at the same frequency, or for the same reasons. A particular subclass of spatial deictics, *demonstratives*, which are apparently so-named for their link to concurrent action, are widely supposed to enjoy a privileged relationship to gesture[51]. Some commentators have gone so far as to argue that the demonstratives *this* and *that* are semantically incomplete without an accompanying gesture (e.g. Levelt et al., 1985). This is an over-statement: demonstratives can be perfectly "complete" without gesture, provided there is sufficiently sturdy extra-linguistic scaffolding for them, as will be elaborated below. Other classes of verbal deictics, such as the person terms *I* and *we*, are generally taken to be semantically complete on their own[52]. Surprisingly, despite a long history of *en passant* recognition, the link between demonstratives and pointing has never been articulated with particular precision. Different dimensions to the relationship have certainly been noted (e.g. Diessel, 2006), but are seldom fleshed out. A first dimension, as just described, concerns semiotic function. Another dimension is that of diachronic origins, the intertwining of verbal and gestural deixis throughout language development and— by speculative extension—

---

[51] Demonstratives are sometimes grouped with spatial deictics, and other times appear to be sequestered into their own special word class (e.g. Deissel, 2006).

[52] Clark (1996) writes that "*I, here, and now* are really demonstrative references for which the accompanying 'gestures' are performed with the voice" (pg. 169). Such a framing obscures the interesting fact that person deictics like *I* and temporal deictics like *now* are not uncommonly co-produced with pointing gestures (see Chapter 3 above; Cooperrider & Núñez, 2009).

throughout language evolution[53]. Of particular concern for the present discussion is a third dimension: their (allegedly) frequent co-occurrence in adult discourse. Demonstratives comprise *entity*-referring forms (in English: *this, that, these, those*), *place*-referring forms (in English: *here, there*), and *manner*-referring forms (in English: *like this, like that, like so, thus*). Entity-referring forms generally appear to be the most frequent in English; the studies reported in Chapter 2 were designed specifically to elicit them, and they are thus the focus of the present discussion[54]. I will now consider the evidence presented in Chapter 2 for four distinct links between pointing gestures and demonstratives.

A first link identified above is that the rate of pointing with co-produced demonstratives hinges on the ability to single out referents unambiguously in gesture. Evidence for this link was seen at both a coarse and a fine grain. At a coarse grain, when pointing by hand, Finders used demonstratives (+pointing) at significantly higher rates in the NEAR condition than in the FAR condition. The best explanation for this difference is that, all other things beings equal, manual pointing gestures become more ambiguous with distance. When pointing by laser, however, there was no difference in rates of demonstratives (+pointing) in the two distance conditions. More fine-grained analyses

---

[53] For example, Bühler (1990 [orig. 1934]) discusses (but does not endorse) what he calls "the myth of the deictic source of representative language" (pg. 101). Such a "myth" has been taken up quite seriously by scholars since, including most recently Tomasello (2008).

[54] Note that, in English, the place-referring forms have distinguishable speaker-*inclusive* and speaker-*exclusive* uses (see Chapter 2, note 12). However, in the studies presented above, speakers only used *here* and *there* in the speaker-exclusive sense, in both cases best glossed as "the place I am pointing to". The present discussion is thus limited to the entity-referring demonstratives and to the speaker-exclusive place-referring demonstratives. Manner-referring demonstratives, though present in the elicited data, are not analyzed in detail.

corroborated the pattern. The lateral distance of each target Fribble from the idealized origo was found to be negatively correlated with demonstrative (+ pointing) rate in the hand-pointing conditions, while the two were uncorrelated in the laser-pointing conditions. Such findings reveal that it is the ability to point out a referent, rather than distance from the speaker *per se,* that shapes the rates at which speakers use demonstratives (+pointing). These findings replicate (and extend by using a modified paradigm and a novel laser pointer manipulation) a result reported previously by Bangerter (2004). Combinations of pointing gestures and demonstratives are designed to uniquely identify referents, and they are resolutely avoided when they cannot be made to refer uniquely. Demonstratives (+pointing) are thus best viewed as a specialized referential device, only deployed given certain contextual affordances.

What, if anything, does this finding imply about the nature of demonstratives? Demonstratives function to focus attention, but they cannot accomplish this on their own. Successful demonstrative reference invariably requires extra-linguistic resources to achieve its specificity. Pointing gestures are one such extra-linguistic resource, and, in the confines of the referential communication task used in the Fribble studies, they are in many ways a privileged resource. The task was deliberately designed to privilege pointing, for instance by neutralizing the perceptual salience of the target creatures (to the extent possible) and by setting the targets far enough away— and at a high enough density— that gaze alone could not be used to fully individuate a target[55]. Other kinds of

[55] Another extra-linguistic resource, which emerged reliably in Study #2 and which is discussed in greater detail below, is the *listener's* attentional focus. The fact that a listener is currently attending or has recently attended to a referent makes certain targets more salient than others.

concurrent action buttress demonstrative reference in different settings. Touching, rapping, displaying, and other sorts of manifest attentional engagement can provide the necessary scaffolding. Clark (2003) has discussed another mode of indicating— *placing*— which he takes to be pervasive (see also Fillmore, 1982 on "presenting"). Pointing and placing, according to Clark, are both means of activating regions of space in perceptual common ground. If a demonstrative cannot be made to refer uniquely by coordinating elements of common ground, it will be avoided in favor of semantically richer referential devices. Finally, and importantly, note that sometimes nothing so overt as pointing or placing is necessary, as in when a referent is presumed to be *de facto* in attentional focus. Take, for example, the case of a sudden loud noise, after which one person says to another "What was *that*?"

A second link presented in Chapter 2 is that pointing gestures are implicated in the proximal-distal contrast, between *this/here* and *that/there*. More specifically, it was found that speakers are more likely to choose proximal forms when they can point unambiguously and distal forms when they are less able to do so. When pointing by laser, Finders used a higher proportion of proximals than distance-matched Finders pointing by hand. It bears emphasis that the task was designed to hold demonstrative-relevant parameters (e.g. familiarity, saliency, distance) constant across the conditions and thus to isolate the *demonstrability parameter*— that is, the ease with which the speaker can project a vector that individuates one of the creatures. It is especially interesting to note that, just as speakers use more demonstratives overall as they are better able to point, they likewise use more *proximals* as they are better able to point. This fact fits with the proposal that the same feature that characterizes demonstrative use generally— *attention-*

directing force— also defines the scale on which the two forms contrast. The finding also appears consistent with a markedness analysis of the proximal-distal distinction, such as that discussed by Levinson (2004). According to a markedness account of the English demonstrative paradigm[56], the proximal demonstrative is the marked term: it contains a special semantic ingredient, an ingredient which in the distal form is attenuated or neutralized. Put more succinctly, the basic idea is that *this* has a more specific and richer meaning than *that*.

A third link identified in Chapter 2 is that pointing gestures are used more frequently with proximal than with distal demonstratives, a pattern that holds across conditions. In both studies, *this* was more strongly associated with +pointing than with -pointing; *that* was more strongly associated with -pointing than with +pointing. A *this*-pointing association as just described was also found in Dutch (Piwek et al., 2008) in the context of a much smaller corpus of demonstrative references (93 in total, see pg. 710). Such a finding fits with the idea presented above of demonstratives being arrayed on a force scale, with *this* being more forceful than *that*. If *this* is marked by a special ingredient of forceful attention-direction, it makes sense that it would be associated with pointing, a behavior functionally specialized for attention-direction; if *that* is marked by neutralization of the attention-directing ingredient, it makes sense that it would be less

[56] Interestingly, the association between the proximal term and marked status does not appear to be universal. Enfield (2003) adopts a markedness–inspired analysis of the Lao demonstrative paradigm, but characterizes the meaning of the distal term as "not here" (marked) while the proximal term has a much more general (unmarked) demonstrative meaning. It should be cautioned, however, as Enfield himself notes, that such analyses of demonstratives suffer from inconsistencies in how markedness status is assigned. Indeed, there is some debate about the utility of the notion of semantic markedness in the first place (Haspelmath, 2006).

strongly associated with pointing. It bears noting that the *this*-pointing association has clear consequences for language comprehension. If *this* is more likely than not associated with an accompanying gesture, and if such a tendency is more or less stable across contexts, comprehenders' expectations and interpretative strategies must adjust. Accordingly, one prediction would be that uses of *this* would trigger an active search on the part of the listener for relevant concurrent action.

The final link identified in Chapter 2, though entirely unanticipated, provides further support for the deictic force account. As has already been noted, pointing gestures are a clear way of manifesting attentional engagement. A perhaps non-obvious prediction made be the force account, then, is that speakers' demonstrative use should be affected, not only by their own pointing behaviors, but *also* by the visible pointing behaviors of listeners. Speakers should adjust how they refer demonstratively to a referent when the listener is concurrently pointing to it. Such an adjustment was observed in Chapter 2 in two ways: in the overall use of demonstrative referring expressions and in the choice of distal forms. First, listeners' manifest attentional engagement provides an extra-linguistic strut, scaffolding demonstrative reference when it would otherwise be ambiguous. Second, speakers overwhelmingly choose the distal demonstrative in such situations. Recall that, according to the force account, distal forms— which reflect not low, but *neutral* force— are motivated by contextual circumstances of different kinds. One such circumstance is the case where high-force would be heavy-handed because the listener is presumed to be already attending to the referent. The near unanimity in choice of a distal form seen in the listener-pointing context— unanimity which stands out sharply against a

backdrop of between-speaker variability— is best explained by the fact that pointing is an unmistakeable sign of attentional engagement.

A number of questions remain about the viability of the force account. One issue is especially vexing, and that is the question of the account's explanatory reach. As presented here, force is meant to account for patterns of demonstrative use in English, but it seems natural to wonder what light it may throw on typologically far-flung demonstrative systems[57]. The applicability of the force account to other two-term systems is straightforward enough, but how might such an account be scaled up to account for three-term systems? According to Deissel (2011) binary demonstrative contrasts are the most common in the world's languages, accounting for 54.3% of languages in his sample, yet a full 37.6% employ a three-way contrast and 8% employ more than three terms. Deissel ventures that multi-term systems show some tendency to be "person-oriented"— that is, rather than encoding only distance from the speaker, they may encode a three-way contrast, for example distinguishing "near speaker", "near hearer", and "far from both speaker and hearer". Much caution is need here, however. Existing reference grammars have seldom lavished attention on the subtleties of demonstrative use (c.f. Enfield, 2003; Levinson, 2004), and to my knowledge there are no agreed-upon elicitation instruments that might conclusively distinguish distance-based from person-based systems (let alone instruments that would stand a chance of disclosing other relevant parameters). It is easy to imagine how an attention-oriented account of demonstratives— such as that on offer

[57] In the literature on demonstratives, there seems to be a widely shared yet seldom defended intuition that demonstratives cross-linguistically rest on a universal semantic foundation, an intuition evident in the startling frequency with which generalizing statements are made on the basis of data from a single language (e.g. Hanks, 2005; Brovold & Grush, *in press*).

here— might be mistaken for either a distance account or a person-oriented account. For instance, when the listener is holding an object, it is expectably in his or her attention; as objects get more distant, they are less likely to be attended; and so on.

In attempting to reconcile the present attention-oriented account with the cross-linguistic prevalence of three-term systems, there seem to be at least two possibilities. A first is that there is simply more cross-linguistic variation in which demonstrative parameters are privileged than is currently appreciated. Attention could rule the day in English and perhaps a handful of other languages; other parameters— distance, here-space, control, etc.— may rule the day in other languages. A second intriguing possibility is that languages could differ in which attentional contrasts are captured by separate lexical items and which are collapsed under single lexical items. In place of the privative, two-term opposition posited in Chapter 2 for English— between strong attention-directing force and neutral force— other paradigms could lexicalize variants of neutral force. As I have argued, neutral force in English may occur because the listener is already attending to the referent, or else because attentional re-orientation is "insecure". But it is perfectly plausible that these two neutral-force variants could be lexicalized separately. How would this look in a three-term system? The former variant might be what is sometimes called "medial"; the latter might be what is sometimes called "distal" (that is, the farthest term in a three-term sequence). In this way, a distinction that is collapsed under a single neutral term in English— *that*— is pulled apart into separate terms. Again, one can readily see how such a three-way set of attentional contrasts might urge an analysis as "person-oriented". Referents under the listener's immediate observation or control are also presumed to be attended; objects neither in the listener's attentional field

nor unambiguously demonstrable might be glossed as remote from the interaction altogether.

The account just described makes at least one non-trivial prediction that could be tested against existing cross-linguistic data. It would predict that, in three-term systems, anaphora would be accomplished by the so-called "medial" term (or, in allegedly person-oriented systems by the "near listener" term). This is because the objects of anaphoric reference are thought to be within the listener's attention. Note that a distance-based account would likely make a different prediction— namely, that anaphora would be accomplished by the "distal" term and cataphora by the "proximal" term (though, in fact, it is not clear from a purely distance-based perspective why it would not be the other way around altogether), leaving the "medial" term altogether out of endophoric duties. This is because endophoric reference is essentially binary in nature, dividing what is past from what is forthcoming; it seems that the most intuitive way to map a ternary distinction to a binary distinction in this case would be to use the two end-points of the scale[58]. Data that could speak to these predictions presumably exist, but have unfortunately not to my knowledge been collated.

5.2.2. Reprise: Morphological variety in pointing

I have claimed more than once above that, at least in North America, speakers operate with a prototype of pointing behaviors, and that one hallmark of this prototype is

[58] This is, of course, just my intuition of what a distance-based account might predict. I am not aware of any on-record speculation about the matter.

the extended index finger[59]. Importantly, the claim is not that index-finger extension is necessarily more frequent than other pointing morphologies— indeed, it is not even clear how such a frequency might be calculated given issues of finding a "representative" sample— only that speakers regard it as somehow distinctive of category membership. I have further claimed that the prototype is misleading. It is misleading in the sense that it distracts empirical attention from the bewildering variety of morphologies actually used for pointing in the wild. The wide world of pointing morphologies includes gestures produced with an open handshape, with the thumb extended, with a fully extended arm, with a flick of the wrist, or with a jerk of the head. As soon as this morphological variety is recognized, it becomes a matter of considerable theoretical interest to determine the factors that condition it. To put a sharper point on the question: why do people point the way they point? When and how do these differences reflect aspects of the discourse context? In this section, in the course of reviewing several parameters of variation in pointing morphology, I propose factors that shape pointing morphology. By outlining such factors more systematically than has been previously attempted, the present discussion will, I hope, pave the way for future studies to specify cross-cultural differences more precisely. What's more, understanding the factors at play in motivating pointing morphology may help lay a foundation for an account of the factors that shape gesture morphology generally.

---

[59] Short of a controlled meta-pragmatic study, different types of informal evidence might be used to support the claim of a pointing prototype. One kind of evidence is that there is a clear etymological link between the act of pointing and the piece of anatomy with which I am claiming it is prototypically associated— the index finger. Another is that a Google image search on "pointing" reveals an overwhelming preponderance of index finger extended handshapes.

Handshape is a first parameter of variation across different instances of pointing gestures. Despite rare anecdotal reports to the contrary (e.g. the case of the Barai of Papua New Guinea, discussed by Wilkins [2003], pg. 176) it seems safe to assume that all human groups, at least occasionally, point with the extended index finger. Biomechanics plays a clear role in this choice. The index finger surpasses the other fingers in its range of motion and strength, and has even been shown to naturally fall to rest somewhat apart from the other fingers (in humans but, interestingly, not in chimps) (Povinelli & Davis, 1994). However, as Wilkins (2003) has argued forcefully, the superficial appearance of a formal universal should not be mistaken for evidence of a *functional* universal. Index-finger pointing across cultures, he argues, varies in how it articulates within a broader pointing repertoire. Heeding this admonition, we can say that differences in pointing handshape across cultures are not just a matter of face-value formal variation, but of differences in how particular handshapes are mapped to particular communicative functions, as well as in which handshapes are understood to be protoypical. To take one example, Arrernte employs a "horned" handshape for pointing, formed by extending the little and index fingers and flexing the middle two fingers back toward the palm. The handshape is used for indicating the global orientation of one's current destination (Wilkins, 2003, pg. 185). Among American speakers such a handshape bears no such conventional associations when used for pointing, though it may well be occasionally used[60].

[60] Multiple-finger handshapes, such as one in which both the index and middle finger were extended, were used— albeit infrequently— in the Fribble studies described in Chapter 2.

Conventionalization processes no doubt explain many observed inter-cultural differences, as pointing handshapes become calcified within certain discursive functions. Explanations of the fluid intra-cultural variation we see in pointing morphology, however, require appeal to other processes. Here I argue that processes of construal are especially important. As discussed in Chapter 3, one explanation for the preponderance of open-hand morphologies in self-points is that they reflect, not morphological noise or a conventionalized form-meaning pairing, but the speaker's *in situ* construal of aspects of the referent. In the case of co-speech self-points, relevant aspects of the target— the speaker's chest— include its size relative to the attentional ground, as well as the part-whole relation holding between the site— again, the chest— and the referent— the whole person. Curiously, related factors appear to motivate morphological contrasts in body-directed signs in ASL. Pyers (2006) notes, for example, that while an index-extended (1-Hand) is used "to label small parts of the body" (pg. 290), a full-hand (Open-B and Bent-B) is used for body parts "spanning a large area" (pg. 292). Such signs are evidently motivated, but are not "pure" fossilized construals. Sign-formation— in contrast to gesture production— is driven by the need to distinguish new signs from existing signs, and this always clouds whatever iconicity they may seem to exhibit.

According to the present account, construal operations evident in self-points are but one instance of a much more general phenomenon in pointing gestures. This fact has received occasional discussion in the literature, but more often only allusion. Haviland (2003) notes that pointing form embodies different kinds of semantic properties "not just direction, but also aspects of shape (or manipulability), and proximity" (pg. 161). Wilkins

(2003) hints more obliquely at construal processes in his discussion of another Arrernte pointing handshape, the "wide hand point". He writes:

> In contrast to the one-finger point, the wide hand point regularly carries with it the notion of non-singularity or non-individuation. It can be used to identify regions or expanses of country (i.e. areas which contain multiple individual places), and is also used to refer to the multiple objects in an area… The orientation of the palm tells the relative orientation of the surface upon which things are extended or spread (pg. 194-5).

What is most interesting about the "wide-hand" point, perhaps, is not its putative place in the quasi-grammatical Arrernte pointing system, but the fact that it seems to be motivated by the same construal processes that motivate morphologically similar pointing gestures elsewhere. The iconicity in Wilkins' example is unmistakeable; it is presumably prior to any conventionalization processes that have served to, as it were, fix the form. It is easy to imagine a process by which recurrent construals over time become entrenched as conventions. Unfortunately, though, we know very little about the dynamics involved. One window into such dynamics may be afforded by examining pointing handshape variation across sign languages and, in particular, across emerging sign languages. A final interesting question raised by a construal account is whether *all* pointing gestures, including the canonical index-extended form, embody construals. After all, the canonical form can readily be seen as embodying a construal of singularity (c.f. Haviland, 2003, pg. 162-3), or, more generally, of narrow focus against a broad attentional ground. Yet it is possible that construal processes motivated the use of the extended index finger initially, but that these subsequently became washed out as the form took hold as a default.

A second parameter of variation in pointing morphology is the choice of articulator used[61]. Current knowledge suggests that all human groups point with their hands, just as it suggests that all human groups at least occasionally also produce non-manual forms of pointing (see McClave et al., [2007] on the universality of deictic use of the head). Cross-cultural differences in articulator selection are thus not absolute, but considerably more subtle, including: first, differences in the frequencies with which certain articulators are used; second, differences in how articulators are specialized for certain communicative functions; and, third, differences in which articulators are considered prototypical pointing instruments by native speakers. Hands and heads appear to be ubiquitously used for pointing, while other articulators— such as the lips or the nose— enjoy much more limited distribution. Still other articulators— the elbow, the foot, etc.— are recruited on an *ad hoc* basis because of their extensibility, but are not currently known to serve as conventional pointing instruments in any cultures. Taking the perspective of human biomechanics, this pattern is hardly surprising. The hands and head are unmatched for the degree to which they are under fine-grained control, as well as the ease with which they can be made visible to an interlocutor.

Evidence suggests, however, that the head and hands are not employed to the same degree as pointing instruments across cultures. This is seen, for starters, in the widespread use of lip-pointing around the world, a form which is basically unknown in North America and Europe. Sherzer (1972) claims, in fact, that lip-pointing is more common among the San Blas Cuna than is manual pointing. In Chapter 4 above, we

[61] In this section I refer to a process of "choosing" or "selecting" pointing morphology. Such words are admittedly inaccurate to they extent that they imply a deliberate planning process when something much more *ad hoc* and unreflective is envisioned here.

suggest that among speakers of Yupno the head has a different status, not just as a pointing instrument, but as a gestural articulator. Evidence for this is found in the frequency with which head gestures are used when the hands are readily available, as well as in the use of the head for both iconic gestures (e.g. tracing a path) and what on a first approximation appear to be discourse-structuring gestures. Complicating such a claim, of course, is the fact that there has been very little systematic research on gestural use of the head, even among Westerners. Judging relative preference for the head or hands in Yupno versus in American speakers is thus a matter of comparing two impressions.

In cultures that make frequent deictic use of both the head and the hands, why do speakers opt for one or the other? While it is possible that articulator choice is unsystematic— again, some kind of morphological noise— a compelling alternative is that such differences in form reflect differences in function. Pursuing this latter kind of explanation, a couple of possibilities exist. A first is that the head is less precise as a pointing instrument than the hand, but at the same time requires less effort to point with. According to a communicative "principle of least effort" (Clark, 1996), speakers should only use hand gestures if a certain degree of precision is required; otherwise they should conserve effort and opt for the good-enough-for-present-purposes head. Precision may be somewhat of a red herring, however: head points, lip-points, and nose-points are all produced with co-directed gaze, and the precision of gaze relative to finger-pointing is in many cases probably comparable (c.f. Enfield, 2001, pg. 208-9, note 7). A second possibility, leaving aside the question of relative precision, it that the choice between hands and head is principally a matter of differential effort. As such, the choice is bound

up with the "robustness gradient" discussed above, from morphologically reduced to morphologically robust, a gradient which is best seen as an outward expression of speaker effort. Head pointing gestures, due to their overall smaller amplitude, may be considered morphologically reduced relative to points involving arm-extension. Speaker effort as manifest in gestural robustness likely reflects non-linguistic motivational factors. But, more interesting for present purposes, is the possibility that it may also reflect properties of ongoing discourse, for example, the information-structural property of focus. Such a mapping between effort and discourse structure is evident prosodically in the phenomenon of lexical stress (Levy & Fowler, 2000), and it may also be evident in gestures accompanying linguistic focus, with robust gestural forms selectively associated with focal aspects of the speech stream[62]. This proposal may provide a generalized account of two recent findings by Enfield and colleagues. First, Enfield (2001) argues that lip-pointing carries a "recognitional flavor", which might be associated with decreased referential effort. Second, Enfield et al. (2007) find that focal location

[62] As seen in Chapter 3, the presence of linguistic focus may condition the very presence or absence of a gesture in the first place. Of course, this cannot mean that all focal words are co-produced with pointing. One hypothesis for future research is that focus status is one privileged factor determining whether a pointing gesture will be "born", while a second privileged factor is the extent to which the referent is associated with a spatial location. Thus high focus elements that are associated with a discrete spatial location will prompt pointing in spades; neutral focus elements that have a discrete spatial might also prompt pointing, but to a lesser degree. At the same time, high focus elements that are not strongly associated with a spatial location may still prompt "abstract" pointing, while elements that are neither high focus nor associated with a particular location will only very rarely prompt pointing gestures. Glimmers of evidence for such a proposal are found in Chapter 3. Self-points were strongly associated with linguistic focus; further, first person singular pronominal words are more strongly associated with pointing gestures than are first person *plural* pronominal words. According to the present hypothesis, this is because *we* is less strongly associated with a discrete spatial location than *I*. Such an account makes a number of predictions at odds with the traditional idea that people (only) point in order to re-orient listeners' visual attention.

reference is associated with robust pointing gestures— "B-points", in their terms— while non-focal reference to location is associated with less robust pointing gestures, or "S-points"[63].

Robustness in pointing is about more than the choice between the head and the hands. Different instances of manual pointing also vary along the robustness gradient. In manual pointing, co-direction of the gaze/ head and the torso are central components of robustness, as is the degree of extension of the arm. Empirical work will be needed to determine which of these components, if any, is privileged over the others, and whether any other parameters might constitute components of robustness. Curiously, aside form the Enfield et al. (2007) paper cited above, the researchers to give the most serious consideration to variation in pointing robustness are comparative psychologists. Miklósi & Soproni (2006), in reviewing work on the comprehension of the human pointing gesture by a range of animal species, are careful to consider the presence of co-directed gaze as a variable meriting experimental control (see Table 1, pg. 84). The researchers consider another important variable to be temporal dynamics, and distinguish between static, dynamic, and momentary pointing gestures. Variation in the temporal dynamics of pointing certainly requires much further attention, whether in observational studies or as independent variables in experimental studies.

Note that a robustness gradient is also at play in *facial* pointing actions. Enfield (2001) hints at such a gradient when he discusses a "full-blown" lip-point, in which different movement features together create the impression of pointing "most

[63] See Chapter 1, note 3 for discussion of why robustness is preferable to Enfield et al.'s (2007) size-based terminology.

emphatically" (pg. 187). In Chapter 4, we described different movement components that define a robustness gradient across different instances of Yupno nose-pointing. These components include the intensity of muscular contraction in the face and the length of time the form is held in its apical position. Any human movement, it seems, can be performed in a way that manifests more or less effort, and exactly how such effort co-articulates with aspects of the speech stream, both in pointing gestures and beyond, is an important research priority going forward.

Above I suggested that the robustness gradient may in part explain the choice between head- and hand-pointing in North America (and presumably elsewhere) and between lip-pointing and hand-pointing in Laos (and presumably elsewhere). The Yupno case shows very clearly the limitations of such an analysis, however. Though smaller in overall amplitude, a nose-point is not properly understood as "less robust" than an index-finger point. This is because, in the Yupno case, the selection of the face for nose-pointing appears to be less a matter of reduced effort than a matter of the particular affordances of the face. Specifically, the face affords the layering over of size information in a way that the hands do not, both because of the scrunching action it can perform and because of the intensity gradient common to all facial actions. From this example we can extract a more general consideration about the cultural selection of pointing articulators: different articulators have different affordances and are matched to particular communicative functions on the basis of these affordances.

Another factor shaping pointing morphology is where the gesture falls in the stream of motor action. All levels of pointing morphology, I propose, are likely subject to effects of "morphological momentum". One finding to come out of Chapter 3 was that

the use of one- or two-hands in self-points is partly contingent on the use of one- or two-handed gestures immediately *preceding* the self-point. From this it can be concluded that self-points are not born *ab ovo* in the moment of production but are built on the fly from available motor resources. This finding is an important proof of concept, but is limited in many ways. It only finds momentum in a highly specific gestural context (self-points), and only along one parameter (use of one or two hands). Investigations of momentum in more contexts and along more dimensions, such as articulator choice or handshape, will be required to understand the generality of the phenomenon. Importantly, though, there is little reason *a priori* to suppose that morphological momentum might be limited to pointing gestures. It would be unsurprising if, in general, two-handed iconic gestures turned out to preferentially give rise to further two-handed gestures, or if right-handed gestures turned out to preferentially give rise to right-handed gestures.

In many ways, the concept of morphological momentum as outlined here echoes McNeill's (2005) notion of a "catchment". He defines a catchment as "a kind of thread of visuospatial imagery that runs through a discourse" (pg. 116-7) and one is recognized "when one or more gesture features occur in at least two (not necessarily consecutive) gestures" (pg. 116). However, McNeill seems to have in mind an explanation for the catchment phenomenon that is very different from the explanation suggested here. He writes: "The logic of the catchment is that discourse themes produce gestures with recurring features; these give rise to the catchment" (pg. 117). What is the source of the observed momentum effects? Are such effects driven from the inside out by "discourse themes"? Or, as argued here, are they the result of mundane motors processes, such as the pressure to select gesture morphology rapidly while also economizing effort? The

question cannot be answered on the basis of the available data, unfortunately. What is at stake in future studies of the phenomenon is of central theoretical interest, however— namely, the nature of the "window" that gesture putatively opens on thought.

5.2.3. Reprise: Lexical affiliates of pointing gestures

Demonstratives are a commonplace and even prototypical accompaniment to pointing gestures, but they are by no means the only verbal accompaniment. What other, semantically richer kinds of speech partner with pointing, and what is the nature of such partnerships? The question has received a somewhat uneven treatment. On the one hand, co-speech pointing has figured prominently in the philosophy of language, thanks to Wittgenstein's and Quine's famous ruminations on the opacity of ostensive reference[64]. On the other hand, the question has received very little empirical attention in recent years. The particular issue to be considered here is the semantic nature of the composites produced in co-speech pointing and the conceptual processes that motivate them. Other issues, such as fine-grained aspects of speech-gesture timing, await future study.

The term "lexical affiliate"— originally coined by Schegloff (1984) and widely used since— refers to the word in the speech stream that bears the closest semantic relationship to the gesture. As an observation about a commonplace pattern in the organization of co-speech gesture, it is unimpeachably important. Many gesture strokes are strikingly co-timed with a single (often stressed) word. As an assertion of a universal pattern, however, it fails, as there are other patterns commonly in evidence (e.g. a gesture

[64] Quine (1960) summarizes the quagmire concisely when he writes: "Point to a rabbit and you have pointed to a stage of a rabbit, to an integral part of a rabbit, to the rabbit fusion, and to where rabbithood is manifested" (pg. 52-53).

filling in after speech falters). Both the utility and limits of the concept are evident in the co-speech pointing gestures considered above. The self-points considered in Chapter 3 most commonly take the form of a short-lived stroke toward the body, tightly co-timed with a first-person pronominal word. Identifying the affiliate in such cases is perfectly straightforward. At other times, as in many of the co-demonstrative pointing gestures described in Chapter 2, the alignment of gesture and speech is much less clear, and the lexical affiliate concept would seem to falter. Complexities of temporal alignment, while evident throughout the Fribble studies, have not been treated above in any detail. Complexities of semantic alignment, however, were of central concern above, and it is these complexities to which I now turn.

In the scant empirical literature on co-speech pointing, a widely shared intuition emerges that certain relations between the targets of pointing gestures and their referents are more common— perhaps more prototypical— than others. The prototype is rarely laid out explicitly, but it is evident in how apparent departures from it are carefully marked by marginalizing terminology. Thus Enfield et al. (2007) devote a section of their discussion to the phenomenon of "secondary pointing gestures". In an analysis of homesign systems, Morford & Goldin-Meadow (1997) review cases of "non-literal pointing". Streeck (2009) discusses a couple of examples of what he calls "indexicality once removed" involving pointing to the body (pg. 143). Borg, after Quine and Nunberg, discusses cases of the puzzle of "deferred ostension" (Borg, 2002), in which a speaker points to some object but in which "what is talked about is not the object indicated" (pg. 490). And, to take a final example, Clark (2005), in considering a similar puzzle, introduces the notion of a "chain of indicating" (pg. 511), implying a backdrop of plain

old unchained indicating. The archetype of co-speech pointing presupposed by such labels would appear to be one in which the *target*— what is pointed to— is present, concrete, perceptible, and perhaps discrete while the *referent*— what is referred to in speech— is some noun or pronoun (whether a proper name, demonstrative, or otherwise) that provides an unassuming, straightforward characterization of the identity of the target. This archetypal scenario would seem to fit many of the pointing examples considered above, such as when a participant in the Fribble studies points to a creature and says "that one", or when Tavis Smiley points to himself and says "I".

Other examples described above however, particularly in Chapter 3, depart from this protoype and so pose analytic challenges. When Tavis points to his chest while saying "we" or "us", he offers— not an unassuming characterization of the target, his chest— but a complex composite reference comprising disparate elements. The TSC includes a colorful mix of spoken referents co-produced with body-directed gestures, including abstract nouns such as "courage", "instinct", and "dream"; adjectives describing felt experience, such as "sick", "comfortable", and "emotional"; and concrete nouns, such as "cancer survivor" and "script". What is striking here is not that there is anything pathological about such speech-gesture partnerships; it is precisely their interpretability despite the complex relations they leverage. Analogous relations in written or spoken language are often considered *metonymic*. To adopt a broad definition, metonymy occurs when a referent is evoked by referring— not to the referent *directly*— but to something else that is associated with it. Metonymy is classically treated as a kind of rhetorical device, one that artfully makes salient particular features of a referent. The metonymies in the present data, however, demand a different treatment for at least two

reasons: first, because they are multimodal, with their complexity straddling speech and gesture; and, second, because they are spontaneously produced.

In analyzing such multimodal metonymies and their motivations, care is required to distinguish a production-side account of the phenomenon from a comprehension-side account. A production-side account asks: What cognitive and/or conceptual processes motivate such utterances? A comprehension-side account asks: What cognitive and/or conceptual processes are involved in interpreting such utterances? One common analytic tack is to effectively blur the distinction between the two accounts by assuming a maximal degree of recipient design[65]. According to such an account, speakers are continually taking pains to make their references interpretable while, at the same time, taking care not to run afoul of Gricean maxims. This kind of account blurs the distinction because the cognitive processes involved in *producing* a reference are presumed to be foremost processes of anticipating how a reference might be *comprehended*. There may be reasons to doubt such an idealization, however. Speakers have been shown to behave more "ego-centrically" than the account suggests, and do so especially under cognitive duress (Keysar et al., 1998). The extent which gesture does or does not reflect recipient design pressures largely remains to be studied empirically. But two relevant findings are worth highlighting: first, the long-standing observation that gesture production is less carefully monitored than speech (McNeill, 1992); second, the fact that gesture rate has been shown to increase with cognitive demands (Melinger & Kita, 2007). Given such

[65] Though they do not consider multimodal metonymies in any detail, this would seem to be the tack taken by Enfield et al. (2007). Their analysis of "S-points " and other so-called "secondary pointing gestures" is framed largely in terms of competing recipient design pressures.

findings, it would hardly be surprising if gesture turned out to be more "egocentric" than speech, at least sometimes[66].

An alternative analytic tack, and the one favored here, is to keep production-side and comprehension-side issues distinct. Accordingly, on the production side, we may note that recipient design pressures are central but are only one kind of cognitive demand on speakers; other includes pressures related to conceptualization, information-packaging, and lexical retrieval. In seeking to explain the production of multimodal metonymies in Chapter 3, I have thus appealed to a suite of conceptual processes: intercorporeality, anchoring, and compression. Of course, the processes invoked to explain the *comprehension* of such cross-modal metonymies may be entirely different (involving, e.g. Gricean inference), and I have had much less to say about these. It should be stressed that the fact that such metonymies are produced on the fly, through a sort of real-time conceptual bricolage, does not mean they are any less interpretable for it. What it may mean is that they owe their interpretability, not to their exquisite designedness, but to the fact that the conceptual processes that generate them are shared, pervasive processes in human reasoning.

Chapter 4, 'Nose-pointing in Yupno', brought to light other issues in the semantic coordination of pointing and speech. According to our provisional analysis, nose-pointing involves a mapping between a very particular pointing morphology— one that, in fact, appears limited to a single cultural group— and a cross-linguistically widespread

---

[66] There is abundant evidence showing that co-speech gesture is communicatively intended and bears the marks of recipient design. Thus the question for future research is not "whether" gesture is recipient-designed, it is more subtle: Are there conditions under which the designedness of gesture is compromised?

grammatical/ semantic category— the diminutive. Links of this kind, between a gestural morphology and a grammatical category, have not been broadly documented. Useful analogues are also somewhat scarce. Kendon (2004) characterizes relationships between pointing morphology and how "the object being referred to is being used in the speaker's discourse" (pg. 201). None of the pairings he describes involve properties of the referent itself, however. Rather, they relate to broader interactive themes, such as whether the speaker is individuating the referent, presenting it for inspection, and so on. Perhaps a more useful analogue is the conventionally contrasting handshapes used in Mayan Mexico to signal the size of different classes of entities, such as humans versus animals (discussed in Haviland, 2003, pg. 141, note). In this case, it would be most interesting to know the extent to which such handshape contrasts track grammatical contrasts.

In exactly what sense is Yupno nose-pointing related to the grammatical category of the diminutive? The relation appears to be one of occasional— but by no means obligate— co-production. That is, the morphological diminutive can be produced independently of the gesture, and the gesture can be produced independently of the morphological diminutive. A potentially revealing model for this kind of gesture-speech relation is perhaps found in the headshake-negation relation in English. Headshakes and verbal negation are occasionally co-produced, but they are also produced independently (Kendon, 2002). Moreover, headshakes are not only associated with explicitly negative words such as "no" and "not" but also with implicitly negative words, such as those marked by the prefix *un-*. This fact perhaps suggests that whatever link exists, it is not between headshakes and a circumscribed class of lexical items, but to a broader conceptual theme. The aptness of this analogy to the study of nose-pointing is largely to

be determined, of course. A major priority for future work is to determine just how closely the gesture "tracks" the morphological diminutive in the Yupno language. Such research could well reveal a looser fit than we currently posit.

5.3. Co-patterning of speech and pointing

At the beginning of this dissertation, I proposed the notion of *co-patterning* as a successor to the notion of co-expressivity. Motivating this proposal, in part, is a simple observation: theoretical frameworks channel empirical attention. Too dogged a focus on relations between gesture and its immediately co-occurring speech may divert attention from broader, theoretically interesting patterns. How gesture relates to individual words in the accompanying speech stream is an interesting question, but it is one of many. A co-patterning approach urges that the immediately co-occurring speech is the starting point for analysis, not the endpoint. The approach zooms out to look for regularities across time, across speakers, across discourse contexts, and across cultures. It tries to understand the processes that generate those regularities, and it looks far and wide for the disparate factors that shape gesture and its partnerships with speech. When confronting a co-speech gesture, the co-patterning approach asks a basic series of "why" questions: why this form, why with this particular speech, and why right now?

By taking such an approach, this dissertation has exposed and explored a number of links between pointing gestures and language. Such links include, among others: the finding that pointing affordances affect whether a speaker uses a demonstrative, as well as which demonstrative form the speaker chooses; the finding that whether a self-directed pointing gesture is "born" in the first place is constrained by information structure; the

finding that rates of self-directed pointing are modulated by the number and case of the co-produced pronoun; and the finding that a highly conventionalized, culture-specific facial gesture in Yupno appears to bear a special relationship to a grammatical/ semantic category in the Yupno language. Each of these links contributes to our understanding of basic questions about when, how, and why people point. And yet, despite these contributions, it must be admitted that, at this point, the co-patterning approach is largely allusive and promissory. Further research will be needed to corroborate the advantages of stepping back from the microscope of co-expressivity.

It should be emphasized that the notion of co-patterning does not imply a directionality of causation. It does not imply, for example, that pointing is under the thumb of language, so to speak, that pointing just reflects structures, categories, and contrasts peculiar to the linguistic system in which it operates. Regularities observed at one level of a phenomenon can co-pattern along with regularities at another level of a phenomenon without either necessarily causing the other. An independent source can simultaneously account for both patterns. And that is often expected to be the case here. This takes us back to the reason that co-patterning is interesting for cognitive scientists in the first place. If we take seriously the suggestion that gesture and speech spring from a common source, as so many have argued, then we should expect manifold evidence of this entwining, and we should expect it across levels of analysis, not any just at any one. In this light, links between pointing and language are an unsurprising— even predictable— outcome of the fact that both spring from a common cognitive source, that both are born of common communicative and conceptual pressures in the flow of real-time thinking and speaking.

Importantly, however, this dissertation has also discussed a number of ways in which pointing gestures are decidedly *un*linked with language— that is, ways in which pointing exhibits regularities that cannot be meaningfully traced to grammar or discourse structure or lexical semantics. The finding of morphological momentum, for example, is interesting in large part because it exposes the *limits* of gesture-speech co-expressivity: it suggests that gesture production is at some level just another kind of motor action and is beholden to humble motor constraints. Likewise, the variety of construal processes in pointing supports very clearly a point that McNeill has argued for more than two decades, the point that gesture morphology is not discrete and segmented like speech but idiosyncratic, continuous, and holistic. Language and language-like conventionalization processes may shape pointing in certain respects, but they do not fully regulate it. Gesture has a life and a mind of its own.

5.4. Conclusion

As others have argued before me, pointing is a "foundational building block" of human interaction (Kita, 2003). But pointing is not a monolith. The scope of variation in pointing— from one moment to the next, from one setting to the next, and from one culture to the next— is wide. I have tried to give the reader some sense of this scope. Understanding the limits of such variation and the regularities underlying it are critical to a scientific account of the phenomenon. I have also tried to show that pointing partners with speech in patterned ways. Pointing thus constitutes an important case study in how language and action are coupled in the course of discourse production. The key question I have tried to answer— or, at least, clear a bit of the path toward answering— is the

following: How do forces like grammar, culture, conceptual structure, and the human body serve to give this building block shape?

References

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*, 415-419.

Borg, E. (2002). Pointing at Jack, talking about Jill: Understanding deferred uses of demonstratives and pronouns. *Mind & Language*, *17*(5), 489-512.

Brovold, A. & Grush, R. (*in press*). Towards an (improved) interdisciplinary investigation of demonstrative reference.

Bühler, K. (1990 [orig. 1934]). *Theory of Language: The representational function of language*. Philadelphia: John Benjamins.

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

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 243-268). Mahwah, NJ: Lawrence Erlbaum.

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse Studies*, *7*(4-5), 507-525. doi:10.1177/1461445605054404

Cooperrider, K., & Núñez, R. E. (2009). Across time, across the body: Transversal temporal gestures. *Gesture*, *9*(2), 181-2006. doi:10.1075/gest.9.2.02coo

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, *17*(4), 463-489. doi:10.1515/COG.2006.015

Diessel, H. (2011). Distance constrasts in demonstratives. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digitial Library, chapter 41. Available online at http://wals.info/chapter/41. Accessed on 2011–09–24.

Enfield, N. J. (2001). “Lip-pointing”: A discussion of form and function with reference to data from Laos. *Gesture*, *1*(2), 185-211. doi:10.1075/gest.1.2.06enf

Enfield, N. J. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, *79*(1), 82-117.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.

Enfield, N. J., Kita, S., & de Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722-1741. doi:10.1016/j.pragma.2007.03.001

Fillmore, C. J. (1982). Towards a Descriptive Framework for Spatial Deixis. In R. J. Jarvella & W. Klein (Eds.), *Speech, Place, and Action: Studies in Deixis and Related Topics* (pp. 31-59). London: John Wiley & Sons.

Hanks, W. F. (2005). Explorations in the Deictic Field. *Current Anthropology*, *46*(2), 191-220. doi:10.1086/427120

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*, 25-70. doi:10.1017/S0022226705003683

Haviland, J. B. (2003). How to point in Zinacantán. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 139-170). Mahwah, NJ: Lawrence Erlbaum.

Kendon, A. (2002). Some uses of the head shake. *Gesture*, *2*, 147-182.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The Egocentric Basis of Language Use: Insights from a Processing Approach. *Current Directions in Psychological Science*, *7*(2), 46-50.

Kita, S. (2003). Pointing: A foundational building block of human communication. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 1-8). Mahwah, NJ: Lawrence Erlbaum.

Levelt, W. J. M., Richardson, G., & Heij, L. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*, 133-164.

Levinson, S. C. (2004). Deixis. In L. Horn (Ed.), *The handbook of pragmatics* (pp. 97 121). Oxford: Blackwell.

Levy, E. T., & Fowler, C. A. (2000). The role of gestures and other graded language forms in the grounding of reference in perception. In D. McNeill (Ed.), *Language and Gesture* (pp. 215-234). Cambridge: Cambridge University Press.

Lyons, J. (1977). *Semantics (2 vols.)*. New York: Cambridge University Press.

McClave, E., Kim, H., Tamer, R., & Mileff, M. (2007). Head movements in the context of speech in Arabic, Bulgarian, Korean, and African-American Vernacular English. *Gesture*, *7*(3), 343-390.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: Chicago University Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: Chicago University Press.

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, *22*(4), 473-500. doi:10.1080/01690960600696916

Miklósi, A., & Soproni, K. (2006). A comparative analysis of animals' understanding of the human pointing gesture. *Animal cognition*, *9*(2), 81-93. doi:10.1007/s10071 005-0008-1

Morford, J. P., & Goldin-Meadow, S. (1997). From here and now to there and then: the development of displaced reference in homesign and English. *Child Development*, *68*(3), 420-35.

Piwek, P., Beun, R.-jan, & Cremers, A. (2008). "Proximal" and "distal" in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, *40*, 694-718. doi:10.1016/j.pragma.2007.05.001

Povinelli, D. J., & Davis, D. R. (1994). Differences between chimpanzees (Pan troglodytes) and humans (Homo sapiens) in the resting state of the index finger: implications for pointing. *Journal of Comparative Psychology*, *108*(2), 134-9.

Pyers, J. E. (2006). Indicating the body: Expression of body part terminology in American Sign Language. *Language Sciences*, *28*(2-3), 280-303. doi:10.1016/j.langsci.2005.11.010

Quine, W. V. (1960). *Word and Object*. Cambridge: MIT Press.

Schegloff, E. A. (1984). On Some Gestures' Relation to Talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 266-298). Cambridge: Cambridge University Press.

Sherzer, J. (1972). Verbal and nonverbal deixis: The pointed lip gesture among the San Blas Cuna. *Language in Society*, *2*, 117-131.

Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*. Philadelphia: John Benjamins.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 171 - 215). Mahwah, NJ: Laurence Erlbaum.