

Gesture-to-Speech Mismatch in the Construction of Problem Solving Insight

J.T.E. Elms
California State University, Northridge

Introduction

Fifty years of research in problem-solving have established that people do not solve problems on the basis of their deep structure, but rather on the basis of their surface form (Hunt 1991, Pinker 2007:272-276, Kirsh 2009). In the case of verbal problems, verbal surface form dramatically outweighs logical problem structure in determining perceived problem difficulty, time to solution, number of errors, and ability to transfer solution experience from one problem to another (Hayes and Simon 1977, Kotovsky et al 1985). This finding motivates a new look at problem solving *as discourse* — focused not on how understood problems are solved, but rather on how problem-solving discourse is understood. Discourse analysis shows promise in this area because people routinely display their cognitive states to each other in speech and bodily signs, fully expecting interlocutors to understand and respond to these displays (Heritage 2005). Moreover, visible actions such as gesture and gaze have been shown to be closely coupled to language production and meaning-making by both speakers and hearers (McNeill 1992, 2005; Goldin-Meadow 2003; Kendon 2004; Streeck 2009).

The present report analyzes a case of problem-solving insight achieved in a dyadic problem-solving discourse task. The task required two participants to work together to solve a murder mystery based on a story by Raymond Chandler. One participant appeared to use propositional speech and gestural simulation as checks on each other while he hypothesized alternative interpretations for the actions of a murder suspect. Each hypothetical scenario began with a gestural metaphor for a named kinship relationship between suspect and murder victim. Irreparable cross-modal mismatch between speech and gesture led to generation of a new kinship metaphor. Merging elements from the two gestural scenarios while maintaining semantic congruence between speech and gesture did not merely precede, but actually constructed, problem-solving insight.

1 The Chandler Murder Mystery Task

To investigate problem-solving discourse in a systematic way, a standardized discourse elicitation task was constructed from a graphic novel adaptation of a short crime story by Raymond Chandler (2003 [1939]). A murder mystery was chosen as the problem domain because murder is an *ill-defined* problem — that is, a problem which cannot be solved by purely logical inference from the information given. In this respect, ill-defined problems contrast sharply with more frequently studied, *well-defined* problems such as mathematical puzzles, chess, Missionaries and Cannibals, or Tower of Hanoi. Because ill-defined problems require frequent redefinition of the problem itself in order to be solved, they are well-suited to studying the construction of meaning in discourse.

The task materials consisted of twelve 7-inch by 11-inch laminated cards, each of which presented scenes from the Chandler story in comic-book style with black-and-white illustrations, balloon dialog, and boxed narration. Each participant was given one scene-of-crime card, one detective-hiring card, and four clue cards. Card assignment and presentation order were randomized. Story elements were distributed across the cards so that each participant possessed half the information available for solving the mystery. Participants were asked to work together to discover what crime was committed, who did it, and why. Participants were not allowed to show each other their cards, but were otherwise free to work with their materials and each other however they wished. The elicited discourse was free-form and all gestures were spontaneous.

The problem-solving session lasted one hour and was video-recorded. Camera angle was chosen to capture the speech, gaze, gesture, and card manipulations of both participants concurrently with a temporal resolution of 1/30th of a second.

The participants in the session shown here were professional problem solvers and coworkers in the technical support group of a mid-sized software company.

2 Insight into a Gesture-to-Speech Mismatch Attractor

One participant, whom I will call TOM, happened to experience a classic instance of problem-solving insight while on camera. (See Davidson 2003 for a review of the literature on problem-solving insight and a description of what it looks like.) The insight event lasted about 22 seconds and occurred in the final minutes of the session while TOM discussed Clue Card 9.¹ To better understand the meaning-making practices associated with this insight event, I excerpted all TOM's utterances involving this card for detailed transcription and microanalysis.

TOM's first mention of the card came 30 minutes into the session and focused on its final three panes (see (1) below). Detective Marlowe is on the phone with his client, a rich old man with an adopted stepson who is also the murder victim.

¹ Cards were labeled for identification with numbers taken from a random number table.

semantically inconsistent with the self-contained and quasi-lexical meaning of a simultaneously occurring, conventionalized hand sign for the number ‘two.’ There is nothing fancy about it.

More indirect forms of mismatch require more analysis to discern because they involve at least one setup step prior to the occurrence of the mismatch event. However, I count these as mismatch as well if semantic inconsistency or contradiction can be demonstrated when the setup step is taken into account.

In TOM’s case, Clue Card 9 turned out to be a powerful mismatch attractor. On three separate passes through this material, he produced three speech errors (two with repair) and three gesture-to-speech mismatches. No other card triggered anything like this cluster of errors and mismatches.

Also intriguing are the reorganizations of person reference that occur in TOM’s successive discussions of this material. Two different kinship relationships are proposed as alternative social identities for the “old man” and the “kid.” TOM then imagines, using both speech and gestural resources, what these kinship relationships might imply for the social dynamics between father and son and the implications for inheritance. TOM ultimately enacts three scenarios:

- *Adopted son scenario* — leads to irreparable mismatch and impasse
- *Stepson alternative* — tested literally “on the other hand”
- *Stepson alternative merged with adoption outcome* — constructs insight

Across these three scenarios, gesture follows a developmental trajectory from well-defined to stylized (distinct iconic handforms become metonymic, stroke amplitudes attenuate, boundaries between spatial index locations becomes less sharply drawn) and from more concrete to more abstract.

3 Adopted Son Scenario: Concrete Metaphor, Irreparable Mismatch

TOM does not construct his adopted son scenario until after he describes the contextual imagery of Clue Card 9 and reads all text aloud to his partner. At this point in the session, both participants have already agreed that the “kid” is indeed the murder victim. He is also a “big blonde bruiser” who “thinks he’s hell with the women” and “drinks a lot, gambles a lot, and loses a lot.” He owes \$50,000 in 1939 dollars (equivalent to \$636,400 in 2010) to his favorite gambling club.

With context established, TOM switches to hypothetical register. In (3) to (5) below, he redefines the “kid” as an *adopted son* and the “old man” as an *adoptive father*. The concurrent gesture is one of scooping something into the lap. It can be glossed metaphorically as simulating a concrete action: ADOPTING A CHILD IS SCOOPING SOMEONE INTO YOUR LAP.

The gesture in (3) to (5) also establishes links between particular regions in space and particular noun phrases — a practice known as *spatial indexing*. This is

Mismatch in Problem Solving Insight

(3) 31:11:11

TOM: did the
RH
LH+1.....+



(4) 31:11:25

TOM: old man adopt
RH $\left\{ \begin{array}{l} \text{3} \\ \text{ } \end{array} \right\} +$

3 Scoop handform arcs up from rest at LoR periphery to MR near torso & down to lap, strikes UR thigh



(5) 31:12:27

TOM: the kid?
RH +.....+



achieved by TOM in two different ways. For the *agentive* noun phrase (NP) “old man” (plus the verb “adopt,” which is included with “man” in the gesture stroke in (4) and might be read as man-who-adopts), the concurrent sweep of the gesture in motion through the region highlighted in yellow establishes this region as indexing the old man. But for the thematic *goal or patient* NP “the kid,” it is the termination of the gesture stroke in the lap immediately prior to spoken utterance of the NP that establishes the region highlighted in purple in (5) as indexing the kid. In short, action in gesture mirrors agency in syntax, while the termination of a gesture at its target mirrors the thematic goal or patient in syntax. This pattern is seen consistently in TOM’s gestures throughout most of the session.

TOM then explores a possible consequence of the adoption: the old man comes into money because of it. The gesture here is an open, palm-up grasping handform in (6) that becomes a palm-up presentation handform with folded thumb, possibly holding something received, in (7). The stroke terminates at a location removed from the body at arm’s length, near the end of the right chair arm. This gesture can be glossed as simulating the action metaphor: COMING INTO MONEY IS GRASPING CASH AT A DISTANCE.

It is the old man who is hypothesized to come into money in speech. Gesture enacts this agentive role by sweeping the grasping stroke through the yellow region previously established as indexing the old man concurrently with utterance of “come into.” Gesture renders the agency of the old man explicitly and visibly, despite the fact that in speech the old man and his thematic role are implied solely by syntax and are never spoken aloud. The parallel use of gesture for “man adopt” in (4) and “[man] come into” in (6) tends to corroborate this interpretation. The gesture stroke in (6) terminates in the green-highlighted region in (7) just before utterance of the goal NP “money”, establishing this region as a spatial index for money after the pattern of thematic role linkage for goal NPs seen in (4) and (5).

In (8), the subordinate clause “because of it” modifies the main clause of (6) and (7), “and come into money.” Although it is fully realized in speech, gesture ignores it completely — there is not even a gesture hold! Instead, gesture skips to the next main clause beginning “and the old—” and completes an enactment for it.

(6) 31:13:01

TOM: >and come into<
 RH +.....↓{4 }+
 LH +1.....+

4 Open grasp palm up becomes presentation handform palm up with folded RT as arm extends to RM front periphery above chair arm



(7) 31:11:25

TOM: >money<
 RH +.....+



Mismatch in Problem Solving Insight

(8) 31:13:2
3



TOM: >because of it? and the old-<
RH +...└(5) }┐

5 Lateral throw starts in RM front periphery above R chair arm, flings across chest to L w/wrist flick, ends palm vertical w/RT & fingers parallel to floor

It would be difficult to justify a gloss for this gesture, were it not for the fact that we have a speech error here that is repaired in the preferred way (Schegloff et al 1977). The follow-on speech is neither independent of the aborted clause nor an elaboration of it that adds new information; it is a repair. As such, it is intended to replace the aborted clause with a minimum of change. From this we can infer the probable content of the original clause aborted in (8), as shown in (9).

- (9) **a. Error:** >and the old-<
b. Repair: and the kid's gambling it away?
c. Inferred Original: and the old [man's gambling it away?]

Now, the stroke in (8) enacts a throwing motion that starts in the money region and flings something laterally away across the chest. Given the content of the aborted speech in (9c), I gloss the gesture as simulating the action metaphor: GAMBLING IS THROWING MONEY AWAY.

In conformance with TOM's usual practice for agentive NPs, the stroke of the gesture passes through the region indexing the old man concurrently with the utterance of "old [man]" in speech. Agency in speech and gesture are semantically aligned: it is the old man who gambles (in the inferred, aborted speech) or throws money away (in the explicit gesture).

However, this assertion contradicts prior cards concerning who is doing the gambling: namely, the kid. This contradiction motivates the repair shown in (10) and (11). And just as speech is minimally repaired to replace "old [man]" with

(10) 31:15:04

TOM: (0.1) and
 RH $\lfloor \text{---}6\text{---} \rfloor$
 LH +1.....+

6 Throw retracts to new start position at UCtr torso, keeps vertical palm handform w/horizontal RT & fingers



(11) 31:15:24

TOM: the kid's
 RH $\lfloor \{7 \text{ ---} \} \rfloor +$
 gambling it away?
 RH +.....+

7 Throw is rotated 90°, sweeps down from UCtr torso to UR thigh in lap



“kid”, so is gesture minimally repaired to modify the agent of the throwing action. The throw is rotated 90°, from horizontal to vertical, so that the stroke can bypass the vertically elongated yellow region indexing the old man. The new downward stroke passes through the purple region that indexes the kid concurrently with utterance of the NP-plus-verb “the kid's,” following TOM’s usual practice for agentive NPs. This correctly aligns agency in gesture with agency in speech.

But the repair introduces a new problem elsewhere in the utterance: **gesture-to-speech mismatch**. It’s all very well to throw money away on the ground rather than off to the side, but if the throw happens to collide with your lap en route, money is not being thrown away after all. The lexical meaning of “away” in speech contradicts the semantics of a throwing gesture that collides with a lap.

When TOM encounters this mismatch, he pauses for 2.5 seconds of what must be construed as think time. But no simple modification can fix the new problem. If he leaves things as they are, he has a gesture-to-speech mismatch on “away.” If he keeps the repaired speech but rotates the gesture back to its original lateral trajectory, any throw with the right hand will pass through the vertically elongated space that indexes the old man. This would erroneously attribute agency for gambling (literally, for throwing) to the old man, and it would also create a new mismatch to “the kid’s” in repaired speech. In short, we have a conflict between two gestural alternatives, *both* of which contradict the repaired speech in different ways. This is irreparable gesture-to-speech mismatch.

There is nothing nonsensical about the repaired speech stream on its own. It is perfectly comprehensible and contradicts nothing in the story. Nevertheless, TOM concludes “that doesn’t make sense” and abandons the topic for 22 minutes.

4 On the Other Hand: The Stepson Alternative

In the adopted son scenario, all gestures were made with the right hand. But in TOM’s second pass through Clue Card 9, all gestures that accompany a hypothetical speech register are enacted literally “on the other hand” — the left. While in this mode, TOM again proposes new social identities for the “kid” and the “old man” — this time as *stepson* and *stepfather*. In the process, he invents a character who is never mentioned in the Chandler story: the old man’s wife.

The gestures that enact the stepson proposal are not representational action simulations, but deictic (or pointing) gestures. Spatial indexes are nevertheless declared in TOM’s usual manner. In (12), the gesture stroke co-occurs with utterance of the agentive NP-plus-verb “he married,” which links the old man to the new region highlighted in yellow. The stroke terminates at TOM’s temple just before utterance of the goal NP “somebody,” which links this somebody — the old man’s wife — to the region highlighted in blue. The stroke in (13) moves through the blue region concurrently with “who had this” in speech, consistent

(12) 54:57:18

TOM: so he married
 RH L{8 }
 LH L—{9 }+

 somebody (0.6)
 RH +.....+
 LH +.....⊥

8 Lays cards face dn on chest
 9 Releases cards, LF1 forms pointer, lifts to L temple



(13) 55:00:07

TOM: who had this ki:d
 LH ⊥{10 }.....⊥

10 Pointer swings forward, LT raised (gun?), stops above L chair arm at arm’s length from body at ML front periphery



with gestural construal of “who” as an agentive NP, and it terminates in the purple region above the end of the left chair arm just prior to utterance of “kid,” after TOM’s usual practice for declaring a spatial index for a goal or patient NP.

It is important to note that these two gestures in sequence move a pointer from Region A (the old man as stepfather) to Region B (the imagined wife) to Region C (the kid as stepson). The manner of movement is unadorned; there is no attempt even to trace a route between regions. TOM is pointing out *connections* with his gestural actions, no more and no less. Crucially, when this sequence is viewed as a composite gesture chain, it *performs an abstracting function* that highlights the indirect nature of the relationship between stepfather and stepson, who are connected only through a third party. (The use of gesture to perform abstractive actions was first described by Streeck 2008:262 and 2009:137, 169-171.)

The use of abstract gesture strokes here does not preclude metaphor. Both post-stroke hold configurations can be construed as attributional metaphors for the linked NPs. In particular, when TOM in (13) points to the purple region above the left chair arm before uttering “kid,” he relocates the kid out of his lap and places him an arm’s length away. This change cannot be accidental, and I would gloss it as: STEPCHILDREN ARE KIN KEPT AN ARM’S LENGTH AWAY.

The gun-like handform used as a pointer in (13) may not be accidental, either, after 55 minutes of rumination on murder — especially when the gun points at the murder victim. It may be that, by raising a thumb on the pre-existing pointer of (12), TOM modulates the base handform with a secondary layer of meaning, to the effect that distant kin may be shot by their relatives in murder stories.

Throughout his enactment of the stepson alternative, TOM’s whole body is skewed to the left, with most of his weight on the left hip, as shown in (14a). But the moment he completes the stepson proposal, he stands and reseats himself with his weight shifted back to the right hip, as shown in (14b). Gesture immediately resumes in the right hand as well and remains there for the rest of the session.

In this right-sided “main track” posture, TOM assesses the stepson alternative against the adoption scenario he abandoned 24 minutes earlier. Simultaneous consideration of the two scenarios is evident in his speech, which quotes the old man

(14) a. Stepson on the Left Hand



b. Adopted Son on the Right Hand



as saying: “°he’s both° my stepson and my adopted so:n.” On “stepson” and “and,” TOM strikes his upper right thigh with his cards at the same spot (in terms of felt impact) that previously indexed the kid as an *adopted* son. These actions prefigure his integration of the two scenarios three minutes later.

5 Constructing Insight: Merging the Stepson and Adoption Scenarios

In the final minutes of the session, TOM returns one last time to Clue Card 9. He first reprises detective Marlowe’s statement about adopting one’s way into money, then is visibly struck by positive affect and switches to hypothetical register as he assembles a solution. Spatially indexed person references from the stepson alternative are merged into the spatial index layout from the adoption scenario. Gesture concurrently traces a causal chain from one merged spatial index to the next in synchrony with speech. In this way, TOM constructs insight into the key plot twist of the Chandler story: namely, that a rich old man might commit murder in order to inherit money from his son.

The first gesture in this sequence establishes the upper right front region highlighted in yellow in (15) as indexing the old man, following TOM’s convention for agentive NPs. (The agentive convention is used to establish spatial indexes for all NPs in this final sequence.) This region is identical to that first assigned to the old man as an *adoptive father* — as is clear from the fact that I’ve overlaid the current gesture in (15) with spatial index highlights from (7) using computer copy-and-paste. Despite an interval of 27 minutes between these two gestures, the fit is exact.

(15) 58:02:04



TOM: the old ma:n: could’ve orchestrated
 RH L{11 }:.....+

11 Vertical palm presentation handform dips & rises in MR front, beats in place

Mismatch in Problem Solving Insight

(18) 58:09:00

TOM: he will (0.3)
RH L{17}:.....↓

17 Downstroke at MR front
near torso, rapid beats



After all, the co-occurring pronoun is lexically ambiguous and syntactically unbound to a referent at this point in the utterance.. Confirmation that reference to the old man is intended in speech does arise in syntax later, though, when “stepson’s” is generated in (21) near the end of the clause. The use of “stepson’s” in that position forces the preceding “he” to refer elsewhere, to the “old man.”

TOM now traces the causal chain to its conclusion. The downstroke on “inherit” in (20) is more pronounced than others and appears to me to metonymically invoke the stroke in (11) that dumped money into the old man’s lap as an icon for inheriting. The grasp in (22) moves the index for money (green) up from its original adoption position, but retains the arm’s-length metaphor for cash at a distance.

Collectively, TOM’s final three strokes perform an abstracting function by connecting the old man to money indirectly, through the stepson. Indirect connectivity makes the stepson an obstacle to wealth for the old man in a way that the adoption scenario, with its direct link between the rich old man and money, did not. In this way, gesture makes a motive for murder both tangible and visible.

TOM exclaims “sakes!” after working this out. He appears to have surprised himself with what his hands have told him. On this evidence, at least, thinking — and meaning — are not always all in your head.

(20) 58:02:04

TOM: inherit
RH L{18 }↓

18 RH lays cards face down
on chest



(21) 58:02:04

TOM: the stepson's
RH ⊥{19 }⊥

19 Presentation handform
tips outward & arcs
out to MR front peri-
phery



(22) 58:02:04

TOM: money.
RH ⊥{20 }...⊥

20 Open hand palm up
reaches up to UR front
periphery



References

Chandler, Raymond. 2003 [1939]. *Trouble Is My Business*. Adapted to graphic novel form by James Rose, Lee Moyer, Alfredo Alcala, and Willie Schubert. In Dean Motter, ed., *Raymond Chandler's Marlowe: A Trilogy of Crime*, New York: iBooks.

Davidson, Janet E. 2003. Insights about Insightful Problem Solving. In Janet E. Davidson and Robert J. Sternberg, eds., *The Psychology of Problem Solving*, 149-175, Cambridge: Cambridge University Press.

Goldin-Meadow, Susan. 2003. *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Harvard/Belknap.

Hayes, John R. and Herbert A. Simon. 1977. Psychological Differences Among Problem Isomorphs. In N.J. Castellan, D.B. Pisoni, and G.R. Potts, eds., *Cognitive Theory*, 21-41, Hillsdale, NJ: Lawrence Erlbaum.

Heritage, John. 2005. Cognition in Discourse. In Hedwig te Molder and Jonathan Potter, eds., *Conversation and Cognition*, 184-202, Cambridge: Cambridge

Mismatch in Problem Solving Insight

University Press.

Hunt, Earl. 1991. Some Comments on the Study of Complexity. In Robert J. Sternberg and Peter A. Frensch, eds., *Complex Problem Solving*, 383-395, Hillsdale, NJ: Lawrence Erlbaum Associates.

Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kirsh, David. 2009. Problem Solving and Situated Cognition. In Philip Robbins and Murat Aydede, eds., *The Cambridge Handbook of Situated Cognition*, 264-306, Cambridge: Cambridge University Press.

Kotovsky, Kenneth, John R. Hayes and Herbert A. Simon. 1985. Why Are Some Problems Hard? Evidence from Tower of Hanoi. *Cognitive Psychology* 17:248-294.

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.

McNeill, David. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.

Pinker, Steven. 2007. *The Stuff of Thought*. New York: Viking.

Schegloff, Emanuel A., Gail Jefferson, and Harvey Sacks. 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. *Language* 53(2):361-382.

Streeck, Jürgen. 2008. Metaphor and Gesture: A View from the Microanalysis of Interaction. In Alan Cienki and Cornelia Müller, eds., *Metaphor and Gesture*, 259-264, Amsterdam: John Benjamins.

Streeck, Jürgen. 2009. *Gesturecraft: The Manufacture of Meaning*. Amsterdam: John Benjamins Publishing.

J.T.E. Elms
California State University, Northridge
2730 Annandale Ln
Simi Valley, CA 93063

jteelms@gmail.com