How two people become a tangram recognition system

Rick Dale
Department of Psychology, The University of Memphis
radale@memphis.edu

Natasha Z. Kirkham
Centre for Brain and Cognitive Development, Birkbeck, University of London
natasha@eyethink.org

Daniel C. Richardson
Cognitive, Perceptual & Brain Sciences, University College London
dcr@eyethink.org

Abstract. In the tangram task, two participants have the same set of abstract shapes, set out in different orders. One participant must instruct the other to arrange their shapes so that the orders match. In the course of the task, participants must find a way to refer to these abstract shapes. In our experiment, we tracked the eye movements of two participants engaged in a computerized version of the task. We found the canonical tangram effect: participants became faster at completing the task from round 1 to round 3. Also, eye-movements synchronize over time. We used cross recurrence analysis to quantify this coordination, and use it to show that as their words coalesce, their actions approximate a single coordinated system.

Introduction

In the current work we ask how a shared reference system transforms the behavior of those using it. Ostensibly, it permits its users to perform reference tasks much more efficiently. If you and I both know what “the jingly one” refers to, each time one of us employs it, the other can sharply orient to the appropriate referent. This skill is most often measured by completion time of these reference tasks. Here we show that something else occurs, more fundamental than simply pace of success: An emerging referential scheme induces partners in a reference task to become a coupled visual attentional system. To do this, we investigate a well-understood joint task used extensively in previous work: the tangram task (see below). Going below previous levels of analysis of word usage and completion times in the tangram task, we quantify the coupling between eye-movement patterns, and show that the signature of attentional coupling changes across rounds as the referential scheme is agreed upon by two task partners. In short, we show the emergence of the tangram-identification system that the pair of participants in this task becomes.

In the tangram task, pairs of participants are asked to identify unfamiliar shapes (Krauss & Weinheimer, 1966; Krauss & Glucksberg, 1969). One of these participants, called the "director," is responsible for describing a chosen shape to the second participant, the "matcher," who does not know which shape but is responsible for selecting it among an array of six possible shapes. The director's and matcher's screens do not match, because the shapes are placed in different orders. The director must therefore use careful description in order for the matcher to succeed. Once all six shapes are identified, they repeat this task. A robust pattern of change occurs as they recognize the six shapes again and again: Participants take less time to solve the task, require fewer words to do so, and end up with a jointly constructed scheme of shorthand descriptions for the shapes (Clark & Wilkes-Gibbs, 1986; see Clark, 1996, Chapter 3, for a detailed review).

Once multiple rounds are performed by the pair, they are capable of effectively identifying tangrams, and completing the task quite rapidly. Indeed, the resultant processes operating in the pair seem to have produced a coherent, functional unit (Hutchins, 1995) of two parts -- a "tangram identification system." This characterization goes beyond metaphor given the constraints of the task and the coordination problem subjects face. An emerging referential scheme weaves director/matcher cognitive processes into systematic tangram recognition performance, and the dyad operates in a fraction of the time that the director and matcher first required to do the task. Here we use gaze tracking and cross recurrence analysis to reveal how two people perform this act of coordination.

Methods

Participants 20 pairs of participants were recruited, and performed the tangram task for class credit. One participant in a pair was randomly assigned to the director role, and the other was assigned to matcher. 8 of these pairs did not provide mouse-movement data do to technical problems. The remaining 12 pairs formed the basis of eye-mouse analyses (see below).

Apparatus Two eye tracking labs on different floors of a building were used. In one of the labs an ASL 504 remote eye tracking camera was positioned at the base of a 17" LCD display. Participants sat unrestrained approximately 30" from the screen. The display subtended a visual angle of approximately 26° x 19°. The camera detected pupil and corneal reflection position from the right eye, and the eye-tracking PC calculated point-of-gaze in terms of coordinates on the stimulus display. A PowerMac G4 received this information at 33ms intervals, and controlled the stimulus presentation and collected looking time data. The second lab used the same apparatus with one difference: The display was a 48" x 36" back projected screen and participants sat 80" away (this lab was designed for infants under a year old). A slightly larger visual angle of approximately 33° x 25° was subtended in this second lab. Participants communicated through hands-free headsets which used an intercom feature on 2.4Ghz wireless phones.

Stimuli 6 tangram shapes were used, similar to those used in previous work. These shapes derive from combinations of common geometric objects (squares, triangles, etc.), and many appear to be humanoid-like forms with subtle distinctions among them. These were projected in a randomized fashion in a 2x3 grid to both director and matcher.

Procedure Once participants obtained their respective director/matcher role (remaining constant throughout the experiment), they proceeded to identify the six shapes in random order. When the matcher identified the 6th (of 6) shape, a new round was initiated by the software, and the tangram shapes were again randomly ordered in the 2x3 display. Three rounds of communication took place.

Data and analysis During the task, at a sampling rate of approximately 30Hz (~33 ms slices), we extracted three behavioral signals: (De) the tangram fixated by the director, (Me) the tangram fixated by the matcher, and (Mm) the tangram fixated by the matcher's mouse cursor. For any given participant pair and communication round, 3 time series were thus produced, two sequences of eye movements and one sequence of mouse movements. For each round, separate analyses were conducted on the 3 possible alignment pairings: director's and matcher's eye movements (De-Me), matcher's mouse and eye movements (Mm-Me), and director's eyes / matcher's mouse (De-Mm). To explore the patterns of coordination in these pairings, we conducted a version of cross-recurrence analysis. This analysis simply compares all time points of two time series, and generates a lag-based percentage of how much matching or "cross-recurring" (i.e., tangram fixation) is taking place at each lag. By plotting this percentage match, known as percentage recurrence or %REC, across all lags, we generate a recurrence lag profile reflecting the pattern of coordination between the two time series (akin to a "categorical" cross-correlation function). When the %REC is largely distributed to the right or left of such a plot, it has direct bearing on the leading/following patterns of the systems producing those time series. For example, consider the darkest line Figure 1a. This is the eye-movement %REC profile for De-Me on round 1, and the largest proportion of recurrent looks is occurring at the lag which indicates the director leads the matcher. This is what would be expected in round 1 (see Dale, Warlaumont, & Richardson, 2011 and Richardson & Dale, 2005 for more methodological detail).

We anticipate that these profiles will change their position and shape as the task unfolds. To quantify how these profiles change, we treat them as distributions of temporal data. For example, in Figure 1a, the darkest line can be treated as a probability distribution of lags. The mean lag will be the central tendency of the overall coordination pattern, kurtosis will reflect how pointed the coordination is, etc. Such a distribution analysis of the recurrence profile will permit us to quantitatively describe how they change shape and position.

We extracted 5 characteristics of the recurrence lag profiles we generated for each dyad, round, and modality combination. First, we measured the overall mean recurrence across the whole profile (avg. %REC). This would be akin to measuring the mean density of a probability distribution (mean of y-axis values). This simply reflects, in a +/- lag window, how much overall cross-recurrence is occurring between two time series. Second, we measured the maximum %REC occurring in the profile. This is equivalent to finding the value of the maximum density (maximum y-axis value), and reflects the maximum recurrence, achieved at one of the lags. Third, kurtosis and dispersion (SD) of the profiles were produced. The first of these measures reflects the pointedness of the coordination. A high kurtosis would indicate the presence of coordination within a small lag window, occurring for a shorter, pointed period of time; lower kurtosis would reflect a broad lag window during which states are recurrent. Dispersion (SD) has the inverse interpretation, and is calculated by treating the profile as a distribution of lags and finding the standard deviation of the sample. Finally, we measured the central tendency (mean) of the profile. This is equivalent to finding the point along the x-axis (here, a lag in seconds) that reflects the center of the distribution. This would measure the overall weighted center of the recurrence profile. A positive or negative mean (different from 0) would be indicative of leading or following by one of the time series.

To our knowledge, this paper serves as a first demonstration of the value of quantifying lag-sequential profiles in this way. We intend these distribution measures to serve as simple yet intuitive quantitative indices that characterize coupled channels.

Results

Completion time. As in previous tangram experiments (see Clark, 1996), dyads become effective at performing the task. Participants required an average of 139.5s in the first round, 58.2s in the second, and 35.6s in the third ($p < .0001$).

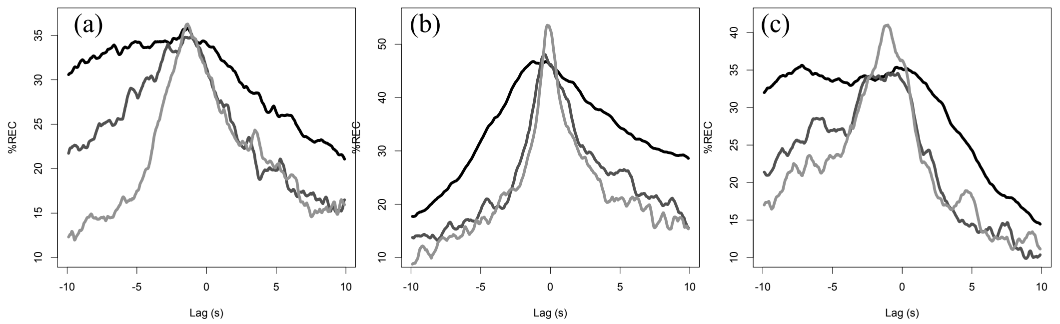

Figure 1. Round 1 is in black, round 2 mid grey and round 3 is in light grey. (a) Direct-matcher eyes, (b) Matcher mouse - matcher eyes; (c) director eyes - matcher mouse

Shuffled vs. non-shuffled lag profile. We first conducted a shuffled baseline analysis for all measures. This was done by performing the same lag-profile analysis but with shuffled versions of our time series. As would be expected, the total recurrence in all analyses within the +/-10-second window we explored was substantially higher in the non-shuffled vs. shuffled conditions (*p*'s < .0001). This main effect of shuffling held in each round when analyzed separately. In short, coordination is significant across all rounds compared to baseline, across all analyses: De-Me, De-Mm, and Mm-Me. The question we explore in distribution analyses below is how that coordination is organized.

Director-matcher eye-movement synchronization (De-Me). The recurrence lag profiles for the alignment between director's eye movements and matcher's eye movements is shown in Figure 1a. To analyze individual distribution values across the 20 pairs, we used a linear mixed-effects model treating subject as a random factor, and tangram round as the sole fixed effect. This analysis revealed several significant changes over rounds.

First, the overall recurrence (mean %REC) drops from round to round, $F(1,57) = 12.2$, $p < .001$, with overall recurrence higher in round 1 (30.3%) than rounds 2 (24.5%) and 3 (21.1%; *p*'s < .005).

Second, there is also a main effect of round for the maximum %REC achieved, $F(1,57) = 3.3$, $p = .04$. Round 1 (39.3%) has a lower maximum %REC value than round 3 (45.0%; p < .05), with round 2 (42.1%) in between (but not significantly differing from these). It is important to note that this maximum difference may not be visible in Figure 1a, because the maximum of the averaged profiles is not necessarily the same as the average of the maximum of the profiles (e.g., consider two non-overlapping normal distributions have higher average maximum, than the maximum of their average).

Third, kurtosis if these distributions increases across rounds, as is indeed visible in the average profiles, $F(1,57) = 14.2$, $p < .001$. Rounds 3 (2.4) and 2 (2.1) had higher kurtosis than round 1 (1.9; *p*'s < .05). Likwise, dispersion in terms of the standard deviation (in seconds) of the profiles is decreasing from round 1 (5.5s) to 2 (5.2s) to 3 (4.8s; $F(1,57) = 20.7$, $p < .001$).

Finally, the mean of this lag profile (in seconds) is changed from round to round, $F(1{,}57) = 7.6$, $p = .001$. The center of these profiles are shifting towards 0s, with round 1 (-.7s) and round 2 (-.8s) significantly lower than 0s, t's < 4, $p < .001$. By round 3, however, the recurrence lag profiles have an average center of .3s, which is not significant from 0, $t(19) = .9$, $p = .4$.

Overall, the recurrence lag profiles between the eye movements of director and matcher is becoming more sharply (higher kurtosis, lower dispersion) synchronous (center near 0) across rounds of communication.

Matcher mouse-movement / matcher eye-movement synchronization (Mm-Me) As noted above under Participants, 8 of the pairs did not supply matcher mouse tracking due to technical errors. We used the time series (Mm and Me) from the remaining twelve to conduct the same linear mixed-effects analyses on the recurrence lag profile characteristics. Parallel to the statistics reported in the previous section, we obtained the following results.

Overall recurrence is again diminishing across rounds 1 to 3 (34%, 24.7%, to 22.3%, respectively), $F(1{,}33) = 10.5$, $p < .001$. Maximum recurrence is not changing over rounds, though the direction of the effect is suggestive of the same pattern (49.8%, 51.3%, and 56.7%, across rounds), $F(1{,}33) = 2.4$, $p = .11$. Indeed, in individual comparisons, round 3 did have significantly higher recurrence than round 1 ($p < .05$), though the overall main effect of round is not significant. Kurtosis did significantly change over rounds, $F(1{,}33) = 4.0$, $p < .05$ (2.1, 2.4, and 2.5 from rounds 1 to 3), though dispersion did not seem to change, but is again in the same direction as seen in the previous analysis (5.1s, 4.8s, and 4.7s), $F(1{,}33) = 1.4$, $p = .25$. The mean of the lag profile did not change, $F(1{,}33) = .11$, $p = .9$. Interestingly however, the mean was stable from round to round (.5s, .6s, .6s) and this mean value was significantly greater than 0, $t(35) = 4.3$, $p < .001$. This suggests that there is a stable leading by the eyes by approximately 590ms.

Though the pattern of significance is different, likely because of the lessened power given lost data, the same general patterns held. The drop in average %REC and increase in kurtosis suggests that the eyes and hand are becoming more sharply coordinate in time. In addition, the stability in the mean value, and significant deviation from 0, suggests a structural limitation of the matcher's hand-eye coordination: there is consistent leading of hand by eye.

Direct eye-movement / matcher mouse-movement synchronization (De-Mm) In analysis of the 12 pairs that provided Mm data, the following results held. First, there appears to be a drop again in mean density of %REC (29.4%, 22.7%, 22.1%), but this is not significant, $F(1{,}33) = 2.1$, $p = .14$. Maximum %REC value is marginally significantly increasing from round to round (42.6%, 48.0%, and 54.6%), $F(1{,}33) = 2.6$, $p = .08$. Kurtosis (2.1, 3.1, and 2.5) and dispersion (5.2s, 4.5s, and 4.6s) also did not achieve significance, $F(1{,}33)$'s < 3. Interestingly, the mean was again relatively stable in these profiles (-1.0s, -1.5s, and -1.0s) indicating that the director's eyes lead the hand of the matcher by approximately 1 second, *t*(35) = -.39, *p* < .001. In general, results argue for an even greater invariant of matcher's hand following the director's eyes, than the delay on the matcher's own eyes.

Conclusion

At the beginning of this task, when director and matcher have not yet become coordinated through referential expressions, the director's eyes lead the matcher's eyes, which in turn lead the matcher's hand. Results suggest that, by the final round, systematic cross-modal coordination has emerged. We characterized this change using distribution analysis over the lag profiles obtain from cross-recurrence analysis. It is not simply that the director and matcher achieve the task faster, but they are strongly synchronized in their shared eye movements. The matcher's hand remains lagged, likely due to an "anchoring" to spatial indices in the visual workspace. As the eyes of director and matcher sample the world to be potentially responded to, the hand stays steady above candidate decisions. With the emerging interplay among multiple behavioral channels, the two participants are therefore acting as a single, coordinated "tangram recognition system."

This characterization of the pair as a single "system" fits well with the backdrop of recent work on coordinating referential domains during interaction. For example, in a review by Tanenhaus and Brown-Schmidt (2008), they showcase extensive recent work in which participants in interactive tasks are subtly influenced by shared and unshared information, suggesting that coordination is a central component of naturalistic interactive tasks. In particular, Brown-Schmidt, Campana, and Tanenhaus (2005) used a complex referential domain to show that attention and comprehension are coordinated tightly as participants get accustomed to the tasks together. In addition, Sebanz and colleagues (Sebanz, Knoblich, & Prinz, 2003) have argued that the very representations and processes used by partners in a task come to overlap simply by being co-present, but particularly by being jointly involved and aware of each other's roles during the task (see also Knoblich & Jordan, 2003). Indeed, the language-as-action tradition (as described in Tanenhaus & Brown-Schmidt, 2008 and Clark, 1996), which sees one person's communication system as largely doing things to or with others, encourages a view consistent with recent perspectives on cognition as "soft-assembling" (e.g., Kugler, Kelso, & Turvey, 1980) into loosely coupled functional systems during interactive tasks (Shockley, Richardson, & Dale, 2009).

The tangram task is a carefully controlled experimental context to measure this soft-assembly of a two-person joint system. The properties of this tangram identification system are highly similar to those that have been identified in individual cognitive systems. We conclude with Hutchins (1995) and Sebanz et al.

(2003) that two-person systems exhibit the same loose coupling under task constraints that a single cognitive processor exhibits, further demonstrating that pairs of people may serve as coherent units of analysis themselves.

References

Brown-Schmidt, S., E. Campana, and M. K. Tanenhaus. 2005. “Real-time reference resolution by naive participants during a task-based unscripted conversation.” In *World-situated language processing: Bridging the language as product and language as action traditions*, eds. J. C. Trueswell and M. K. Tanenhaus. MIT Press, p. 153–171.

Clark, H. H. 1996. *Using language*. Cambridge University Press.

Clark, H. H, and D. Wilkes-Gibbs. 1986. “Referring as a collaborative process.” *Cognition* 22(1): 1–39.

Dale, R., A. S Warlaumont, and D. C Richardson. 2011. “Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams.” *International Journal of Bifurcation and Chaos* 4.

Hutchins, E. 1995. “How a cockpit remembers its speeds.” *Cognitive science* 19(3): 265–288.

Knoblich, G., and J. S Jordan. 2003. “Action coordination in groups and individuals: Learning anticipatory control.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29(5): 1006-1016.

Krauss, R. M, and S. Glucksberg. 1969. “The development of communication: Competence as a function of age.” *Child Development* 40(1): 255–266.

Krauss, Robert M., and Sidney Weinheimer. 1966. “Concurrent feedback, confirmation, and the encoding of referents in verbal communication.” *Journal of Personality and Social Psychology* 4(3): 343-346.

Kugler, P. N., J. A. S. Kelso, and M. T. Turvey. 1980. “On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence.” *Tutorials in motor behavior*: 3-47.

Richardson, D. C, and R. Dale. 2005. “Looking To Understand: The Coupling Between Speakers’ and Listeners’ Eye Movements and Its Relationship to Discourse Comprehension.” *Cognitive Science* 29(6): 1045–1060.

Sebanz, N., G. Knoblich, and W. Prinz. 2003. “Representing others’ actions: just like one’s own?” *Cognition* 88(3): 11–21.

Shockley, K., D. C Richardson, and R. Dale. 2009. “Conversation and coordinative structures.” *Topics in Cognitive Science* 1(2): 305–319.

Tanenhaus, M. K, and S. Brown-Schmidt. 2008. “Language processing in the natural world.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1493): 1105.