

"Conversational" Coupling of Gaze Behavior in Prelinguistic Human Development

Joseph Jaffe,¹ Daniel N. Stern,¹ and J. Craig Peery¹

Received September 1, 1972

Mathematical regularities in the gross temporal pattern of infant-adult gaze behavior are identical to those found in adult verbal conversations. Both types of interaction conform to a Markov chain model. Such regularities suggest some universal property of human communication which predates the onset of speech. The present infants were 3.5 months old.

INTRODUCTION

In the third or fourth month of life, the normal human infant enters the babbling stage of prelinguistic vocalization. Yet by adult standards the actual incidence of vocalization at this age is extremely low. In fact, when physical separation requires that social interaction be conducted solely via the distance receptors, the infant's contribution consists primarily of "looking." By 3.5 months, the visual-motor system is functionally mature, permitting fine voluntary control of gaze (White *et al.*, 1964), and this characteristic uniquely qualifies gaze behavior for the regulation of prelinguistic social contact. Moreover, the infant's linguistic environment is never a disembodied voice. It has even been suggested that vocalizations produced beyond the vision of an

This research was supported by the Grant Foundation, the Research Foundation for Mental Hygiene, Inc., the New York State Department of Mental Hygiene, and the Federal Work-Study program of the City College of New York.

¹Department of Psychiatry, College of Physicians and Surgeons, Columbia University, and New York State Psychiatric Institute, New York, New York.

infant may not be immediately relevant (Vetter and Howell, 1971). For all normal infants, then, the so-called linguistic environment is inextricably interwoven with facial expression and eye contact.

The natural experiments of total deafness and blindness provide dramatic demonstrations of the redundancy and potential flexibility of the intact system. Although vision is obviously necessary for the normal utilization of kinesic paralanguage, blind infants can ultimately develop normal speech if given sufficient auditory and tactile "dialogic" experience (Fraiberg, 1971). In contrast, audition is absolutely critical for the emergence of normal speech. These experiments of nature inform the understanding of intact systems by defining their limits but cannot substitute for studies of optimal situations.

In normal development, when all sensory modalities operate synesthetically, a highly articulated "conversational" pattern characterizes the gaze behavior of the babbling period. This prelinguistic kinesic interaction is therefore of potential theoretical significance for the problem of uncomplicated language acquisition.

METHOD

Subjects

The subjects were six normal infants (two male and one female sets of twins). In the present report, we treat these twins as six separate individuals; clinical, genetic, and developmental implications are discussed elsewhere (Stern, 1971; Stern, 1973).

Procedure

Each infant was observed in the fourth month of life, during free play with its mother and with several experimenters. Physical separation between adult and infant was 1-3 feet. Otherwise, naturalistic conditions were sought and adhered to as much as possible. That is, observations were conducted in the home setting, the adults were instructed simply to behave normally, and the play period was terminated whenever the adult felt it had run its natural course. The interactions were consequently of variable duration, and no attempt was made to standardize these durations artificially. All free play sessions occurring within a week of the infants' fourteenth week of life were analyzed. This yielded 27 separate play periods, 14 between infants and mothers and 13 between infants and experimenters.

Interactions were recorded on an Esterline-Angus event recorder at a paper speed of 15 cm/min. One observer was positioned several yards behind the infant and scored, by means of a pushbutton, adult visual fixations of the infant's face. Another observer was positioned behind the adult and similarly scored infant fixations of the adult's face. The observers worked simultaneously, each activating separate channels of the recorder. The dyadic gazing record so obtained was sampled at 0.6-sec intervals, each sample being coded into one of the following four dyadic states: (1) neither looking, (2) only the infant looking, (3) only the adult looking, and (4) both looking at one another. When adjacent states were found to be identical, i.e., when a state followed itself, it was assumed to have been continuous between samples. The sequences of states was then analyzed as a stochastic process.

Analysis and Results

For each play period, a transition matrix, P , was computed, the elements of which, p_{jk} , are the conditional probabilities of moving from state j at time $t-1$ to state k at t ($j, k = 1, 2, 3, 4$). One such matrix is shown in Table I(b). It is a stochastic matrix the rows of which each sum to unity. Each row of the matrix is a probability distribution of next states given the prior state indicated by the row.

Our first question is whether of 0.6-sec intervals yield statistically independent samples of the dyadic pattern. Inspection of Table I(b) is

Table I. Illustration of the Dyadic Coupling Mechanism^a

I(a)				I(b)			
30	11	3	0	.68	.25	.07	.00
8	58	0	14	.10	.73	.00	.18
6	0	30	24	.10	.00	.50	.40
1	11	36	324	.00	.03	.07	.90
I(c)							
.70	.23	.05	.02				
.08	.74	.02	.16				
.06	.04	.54	.36				
.00	.03	.07	.90				

^aObserved matrix of transition frequencies.

^bStochastic matrix of transition probabilities, generated when each element of (a) is divided by its row sum (rounded to 2 places).

^cIndependent decision matrix, generated from (b) according to equation (1). Comparison of corresponding elements of matrices (b) and (c) shows the discrepancy between the data and the model. The discrepancy is least for those elements which are based on high frequencies as shown in matrix (a).

sufficient to reject this possibility. If adjacent samples were statistically independent, the probability distributions for all four rows of the matrix would be identical; i.e., the transition probabilities into any state k would be the same regardless of the preceding state j . Since the values of p_{jk} obviously depended on j , a Markov chain was proposed as the simplest model of sequential constraints in the time series. A finite Markov chain is a stochastic process which moves through a finite number of states, and for which the probability of entering a certain state depends *only* on the last state occupied (Kemeny and Snell, 1960). Stationarity was assumed; i.e., the hypothesis that the transition probabilities are constant throughout the play period was not tested. The null hypothesis that the chain is at most first order (that only a 0.6-sec history is operative) was tested against the alternative that it is at least second order (that a 1.2-sec history contributes significant additional information). The appropriate test for this hypothesis is the likelihood ratio criterion which is described in the Appendix. As shown in Table II, the null hypothesis was rejected in five out of the 27 play periods at the 0.05 level of significance, in three of these five at the 0.01 level, and in one of these three at the 0.001 level. Only one or two rejections would be expected by chance, so the five deviant interactions were examined further. Table II indicates that the first-order model fits at least once for each of the six infants studied, that the deviations occur in three of the six infants' interactions and are not systematically related to mothers or experimenters. Since the statistic is sensitive to the total number of observations (N), the variable durations of the play periods are a possible explanation for the poor fits. Table II shows that three of the exceptions were indeed among the longest play periods but that the two others, including the worst fit, were shorter than the average. Furthermore, the 27 χ^2 values were not significantly correlated ($r = 0.334$) with their associated numbers of observations. Finally, it is of interest that the expected median value of χ^2 for 36 degrees of freedom is 37, yet only seven of our play periods fall above this value whereas 20 fall below.

From these statistical considerations, we conclude (1) that the overwhelming majority of the play periods fit the first-order Markov model, implying that only a 600-msec history is operative in each transition, and (2) that the five deviant interactions represent genuine rather than chance departures from the model and require further explanation. For example, the transition probabilities are averaged over the entire play period. Yet we have detected important but infrequent subsequences which momentarily impose a special interactive unit on the overall sequence. For example, a common "game" between mother and infant is one in which the baby repeatedly turns playfully toward and away from the mother while she stares at him fixedly. This yields a 3-4-3-4-3-4 . . . sequence of dyadic states. The converse "game"

Table II. Test of Hypothesis That Markov Chain Is First Order^a

With mother			With experimenter		
Infant	<i>N</i>	-2 ln λ	Infant	<i>N</i>	-2 ln λ
A1	819	23.28	A1	546	35.47
			A1	219	17.36
			A1	1110	51.86 ^b
A2	613	25.65	A2	776	55.25 ^b
			A2	463	58.27 ^c
B1	837	59.94 ^c	B1	415	27.54
B1	438	69.59 ^d	B1	438	7.97
B1	403	12.63			
B2	611	33.21	B2	625	35.21
B2	220	42.96	B2	707	21.97
B2	442	6.80			
C1	409	21.79	C1	721	42.53
C1	473	11.27	C1	575	16.16
C1	547	13.48	C1		
C2	331	26.27	C2	501	34.15
C2	778	21.30	C2	582	15.27
C2	404	7.26			

^aThe twin pairs are labeled by letter and the individual infants in each pair by number. *N* is the number of 0.6-sec samples per play period.

^bSignificance $p < 0.05$ ($df = 36$).

^cSignificance $p < 0.01$ ($df = 36$).

^dSignificance $p < 0.001$ ($df = 36$).

also may occur, namely, a 2-4-2-4-2-4... sequence in which the mother playfully turns toward and away from the staring infant in a variant of the "peek-a-boo" game. State 4, which appears in both games, is intuitively a different event in the two examples and may prove to exhibit two different distributions. This would require more complex models, perhaps of the "projected" Markov chain variety, as found necessary in analysis of mature spoken conversations (Jaffe, 1970; Anderson and Jaffe, 1971).

MECHANISM FOR THE DYADIC COUPLING

The notion of dyadic states is not intuitively satisfying. An interactive hypothesis which accounts for these joint configurations in terms of the

individual behaviors of the participants would be more intelligible. We therefore propose a model of interpersonal constraint (Jaffe and Norman, 1964) which has been widely applied to gross temporal patterns of vocalization in mature conversations (Jaffe *et al.*, 1967a,b; Jaffe, 1968, 1970; Jaffe and Feldstein, 1970). The model assumes that both mother and infant make statistically *independent* decisions to look-at or look-away in each sampling instant. The outcome of the pair of independent decisions is one of the four dyadic states. However, each of these individual decisions is contingent on the *joint* looking state which was so generated in the previous 0.6 sec. This contingency imposes both sequential and interpersonal constraint.

$$\begin{aligned} \text{Let } q_i &= \text{Pr}(\text{baby looks at mother at time } t \text{ given dyadic state } i \text{ at } t-1) \\ r_i &= \text{Pr}(\text{mother looks at baby at time } t \text{ given dyadic state } i \text{ at } t-1) \\ 1-q_i &= \text{Pr}(\text{baby looks away at time } t \text{ given dyadic state } i \text{ at } t-1) \\ 1-r_i &= \text{Pr}(\text{mother looks away at time } t \text{ given dyadic state } i \text{ at } t-1) \\ & \text{(where } i = 1,2,3,4 \text{ and Pr denotes probability)} \end{aligned}$$

The method for estimating these parameters from matrix P is given in the Appendix. The four-state Markov chain generated from these assumptions hypothesizes no decision lasting longer than 600 msec. The transition matrix, \hat{P} , for this interactive model is shown in equation (1) and the prediction for one play session in Table 1(c).

$$\hat{P} = \begin{bmatrix} (1-q_1)(1-r_1) & q_1(1-r_1) & (1-q_1)r_1 & q_1r_1 \\ (1-q_2)(1-r_2) & q_2(1-r_2) & (1-q_2)r_2 & q_2r_2 \\ (1-q_3)(1-r_3) & q_3(1-r_3) & (1-q_3)r_3 & q_3r_3 \\ (1-q_4)(1-r_4) & q_4(1-r_4) & (1-q_4)r_4 & q_4r_4 \end{bmatrix} \quad (1)$$

Note that \hat{P} is also a stochastic matrix the rows of which each sum to unity. As seen in Table 1, it is a reasonable hypothesis regarding the coupling mechanism responsible for matrix P , accomplishing the linkage by means of four parameters for baby and four parameters for mother. It is of interest that the empirical values of the q_i parameters indicate that the baby indeed discriminates the four dyadic states of the interaction. That is, if $q_1 = q_3$ and $q_2 = q_4$ it would mean that the infant was operating in only two rather than four states and was not responding differentially to the mother's looking behavior. This observation is important methodologically, since it provides a quantitative tool for discovering the onset of this discrimination in babies younger than 3 1/2 months. It may also identify a disengagement in an infant (or mother) who already possesses the discrimination. Such disengagement is commonly described as "looking through" the target, gazing into the "middle distance," "tuning out," or simply staring.

RELATION TO TEMPORAL PATTERNS OF SPOKEN CONVERSATION

The foregoing model of prelinguistic kinesic interaction assumes additional interest in light of analogous chronographic studies of another dyadic system, namely, the four states of adult conversation: (1) neither speaking, (2) only person A speaking, (3) only person B speaking, and (4) both speaking simultaneously. The choice of a 0.3-sec sampling interval for the conversational analysis was dictated by linguistic considerations such as syllable and juncture pause length. However, as a first approximation, a four-state Markov chain was a useful model for this simplistic coding of mature linguistic interaction. Although the model discards all sequential information longer than 300 msec, its purely mathematical consequences yielded accurate estimates of relatively long (2-10 sec) conversational events such as the average length of time a speaker "holds the floor" (Jaffe *et al.*, 1967a). These estimates were just about as accurate under the hypothesized coupling mechanism of equation (1) (Jaffe *et al.*, 1967b).

DISCUSSION

Two complex dyadic communication systems have each been described in the same grossly simplified form. The simplification consists in the coding of each participant's behavior in terms of only two states, i.e., looking-at or looking-away for one system and vocalizing or pausing for the other. Thus the information encoded for each system is limited to four joint configurations. The temporal sequence of the four states is then digitized and analyzed as a stochastic process. A Markov chain structure is discovered for both systems such that the probability of entering any dyadic state usually depends on a history of less than a second. The implication of the finding is that for both systems the sampling interval is shorter than some unit, with the result that sequential constraint is observed between samples, and that the probability of remaining in each state is relatively constant. Discovery of a Markovian structure, so ubiquitous in nature, as common to the two systems might be merely a mathematical curiosity were it not for the fact that these systems are so intimately coordinated in normal conversation. We do not consider the "conversational" coupling of early gazing behavior to be a precursor of verbal conversational patterns—it is rather, more obviously, a precursor of later gazing patterns. However, both are communicative systems whose maturation requires coordination with the other. Our central point is that the common mathematical regularity may represent some universal formal property of dyadic communication which is detectable in infant gazing behavior long before the onset of speech.

APPENDIX

The test of the null hypothesis involves computation of transition probabilities with a two-step history. These are compared to the one-step transition probabilities shown in Table I(b) to see if the additional history contributes significant information, i.e., increases the probability of the observed sequence. The appropriate test is the likelihood ratio (λ) criterion (Anderson and Goodman, 1957). We compute $-2 \ln \lambda$, which has an asymptotic χ^2 distribution,

$$-2 \ln \lambda = -2 \sum_{ijk} f_{ijk} \ln (p_{jk}/p_{ijk})$$

where p_{ijk} is the probability of being in state k at time t , having been in state j at $t-1$ and in state i at $t-2$; f_{ijk} is the frequency of each triplet type; and $i, j, k = 1, 2, 3, 4$.

In terms of matrix P , maximum likelihood estimates of the coupling parameters are

$$q_i = p_{i2} + p_{i4}; r_i = p_{i3} + p_{i4}; i = 1, 2, 3, 4$$

ACKNOWLEDGMENTS

We thank Mr. Edward M. Lorick for computer processing of the data, Dr. Joseph Fleiss for statistical consultation, and Dr. Samuel W. Anderson for important theoretical criticisms.

REFERENCES

- Anderson, S. W., and Jaffe, J. (1971). Projections between memory codes and some speech events not derivable from stimulus-response theory. *Math. Biosci.* 12:303-320.
- Anderson, T. W., and Goodman, L. A. (1957). Statistical inference about Markov chains. *Ann. Math. Stat.* 28:89-110.
- Fraiberg, S. (1971). Intervention in infancy: A program for blind babies. *J. Am. Acad. Child Psychiat.* 10:381-405.
- Jaffe, J. (1968). Computer assessment of dyadic interaction rules from chronographic data. In Shlein, J. (ed.), *Research in Psychotherapy*, Vol. 3, American Psychological Association, Washington, D.C., pp. 260-276.
- Jaffe, J. (1970). Linked probabilistic finite automata: A model for the temporal interaction of speakers. *Math. Biosci.* 7:191-204.
- Jaffe, J., and Feldstein, S. (1970). *Rhythms of Dialogue*, Academic Press, New York.
- Jaffe, J., and Norman, D. A. (1964). A Simulation of the Time Patterns of Dialogue, Scientific Report No. CS-4, Center for Cognitive Studies, Harvard University, Cambridge, Mass.

- Jaffe, J., Feldstein, S., and Cassotta, L. (1967a). A stochastic model of speaker switching in natural dialogue. In Salzinger, K., and Salzinger, S. (eds.), *Research in Verbal Behavior and Some Neurophysiological Implications*, Academic Press, New York, pp. 281-294.
- Jaffe, J., Feldstein, S., and Cassotta, L. (1967b). Markovian models of dialogic time patterns. *Nature* 216:93-94.
- Kemeny, J. C., and Snell, J. L. (1960). *Finite Markov Chains*, Van Nostrand, Princeton, N.J.
- Stern, D. N. (1971). A micro-analysis of mother-infant interaction: Behavior regulating social contact between a mother and her 3 1/2-month-old twins. *J. Am. Acad. Child Psychiat.* 10:501-17.
- Stern, D. N. (1973). Mother and infant at play: The dyadic interaction involving facial, vocal and gaze behaviors. In Lewis, M., and Rosenblum, L. (eds.), *The Origins of Behavior*, Vol. I, Wiley, New York.
- Vetter, H. J., and Howell, R. W. (1971). Theories of language acquisition. *J. Psycholing. Res.* 1:31-64.
- White, B. L., Castle, P., and Held, R. (1964). Observations on the development of visually-directed reaching. *Child Dev.* 35, 349-364.