Minds and Cultures: Particle Swarm Implications

James Kennedy

Bureau of Labor Statistics
Kennedy_Jim@bls.gov

Particle swarm theory suggests that minds and cultures are both effects of local social interaction. This paper compares two social adaptation algorithms and proposes a view of intelligence or cognitive adaptation as immerging from culture, which emerges from social interaction. A framework for the depiction of mental states is presented, and the optimizing effect of social interaction is demonstrated in a parallel constraint satisfaction paradigm.

My argument in this paper will be that cultures and minds are global and local aspects of a process which can be described in terms of social interaction. I intend to draw a link between the emergence of norms and culture from local interaction on one hand, and the *immergence* of intelligence from culture on the other.

Social Influence

A very large amount of social-psychological research demonstrates that people are influenced by the people around them, in their attitudes, their beliefs, their behaviors, and other aspects of the way they process information. People who interact more frequently come to have more in common, and converge on cognitive patterns which distinguish them. While we tend to think of conformity as a negative trait, social influence is a constant presence. For instance, we all speak the language of our communities, though infinitely many possible languages exist, and a great many actual languages do exist. We don't choose a language, we conform to our social group.

A salient example of beneficial social influence is seen in Kuhn's (1970) description of the process of science. Kuhn shifted the locus of scientific discovery away from the individual, emphasizing the emergence of "normal science" and "paradigms" comprising numbers of scholars working within a shared theoretical and methodological framework. Kuhn's insight suggests that conformity, social influence, and group processes are powerful media for the development of innovation, insight, and wisdom – social influence is not just the vanity of teenagers' fussing over the latest fashions, though the social-psychological mechanism may be the same.

The truth of a statement can be determined by two kinds of criteria: empirical observation and deductive certainty. Yet it is widely recognized that neither of these criteria can generally be satisfied in the real world; except under rare and rigorous laboratory conditions we are usually not privileged to witness the unconfounded working of one variable upon another, and only in the rarefied context of mathematics is deduction infallible. Therefore we social beings rely on a less rigorous criterion for establishing the truth of statements: *agreement*. If others around us believe something to be true, especially when they appear knowledgeable, then the statement can attain the status of veracity. The current paper will suggest that the method of agreement is a highly effective method for discovering truths in complex situations.

Intelligence and Minds

Sternberg (1988) wrote, "Intelligence is essentially a cultural invention to account for the fact that some people are able to succeed in their environment better than others" (p. 71). I have further argued (Kennedy, 1996) that intelligence is simply a term that stands for qualities of a "good mind." As such, the concept is fundamentally a judgment, and, because we consider ourselves to be good people, qualities attributed to intelligence are likely to be autobiographical in nature. In studying real minds and creating artificial ones, our definitions of intelligence as "good mind" are bound to be influenced by our ideals about ourselves and our culture's norms, as well as by the task that the mind is concerned with. We would like for our creations to do what we would want ourselves to do.

Attribution theory is a social-psychological approach to the study of how people assign causality to events, especially events involving persons (cf. Kelley, 1973). This perspective generally focuses on individuals' inferences that behaviors have internal versus external, as well as stable versus unstable, causes. We may infer that a person did something because of a more-or-less permanent personality disposition, because of a temporary mood or physical state, or because the situation required it.

The recurrent finding of attribution theorists is a propensity to exaggerate our autonomy as individuals. In social psychology, the "fundamental attribution error" (Ross, 1977) is the tendency to assign greater causal responsibility for events to persons rather than to situations. We underestimate the causal power of situations and overestimate the power of people – we are more likely to say, "He acted that way because that's the kind of person he is," than, "The situation forced him to act that way." As a result, when we consider how to make computers intelligent – that is, how to make them act like ourselves – we tend to start with the assumption that intelligence is a trait of the individual; we commit the fundamental attribution error. The result is a focus on "autonomous" agents, even though we ourselves are hardly autonomous at all!

I suggest that an aspect of the fundamental attribution error is the belief that minds exist inside persons, and that information processing occurs entirely in private. One needs only to consider feral humans (Lane, 1976), individuals who have grown up without social interaction, to see that *without social interaction a person would not have anything that could be called a mind.* We think by interacting with others, "actual, imagined, or implied" (Allport, 1985); the communicative aspect of language is inseparable from its representational aspect. Our minds may be experienced phenomenologically as isolated entities inside hermetically sealed cognitive chambers, but this is an illusion – a disconnected mind would be no mind at all.

If we as scientists are trying to elicit intelligent behavior from a computer, and if intelligence can be (as I am taking the liberty) defined as an idealized self-portrait, then we would do well to program the computer to do what we do – and what we do more than anything else is interact with one another. Interacting, we become more alike, we conform. And conforming, we converge on good ways to behave: we become intelligent.

Social Adaptation of Knowledge

The exposition of the social model of mind starts with the articulation of two assumptions that are implicit in any theory of thinking. First, a model assumes that the state of a mind can be depicted in some way. For instance, a (crisp- or fuzzy-) logical depiction of a mind might contain a list of statements and the relations among them; a constraint-satisfaction depiction might be a graph with nodes representing beliefs, and connections representing constraints; a neural network depiction might show a mind as a set of connection weights. There is no "correct" or even best method for depicting all minds at all points in time. "...even within European-American culture, theory-of-mind content is not as consistent as the literature might lead one to expect. What is held forth in academics as *the* theory of mind is actually a European-American formulation, one that resonates with scientifically-minded academics" (Lillard, 1997, p. 268). On the other hand, a depiction of the relevant qualities of the state of a particular mind at a particular time, seen from a particular theoretical perspective, is entirely possible — these depictions are useful and even common.

The second assumption is that the state of a mind can be evaluated. A logical mind can be evaluated in terms of the certainty of its conclusions; a constraint-satisfaction mind can be evaluated in terms of how well the constraints are satisfied, and so on. When we talk about "states" of mind, we presume that minds change over time; the measurement of the goodness of a state of mind is not a judgment of character, but rather an assessment of the goodness of the selected qualities of a mind at a given moment. Every theory of scientific or folk psychology has a way of evaluating the goodness of a mind; a mind's intelligence then can be defined within that framework as the goodness of its states averaged over a number of critical measurements.

The state of a mind can be represented by some vector of numbers x_d. As above, these numbers might be neural net weights, attitudes, set memberships, beliefs, activation values, etc. Further, every vector $\mathbf{x}$ can be evaluated by some goodness function $G(\mathbf{x})$, which is determined by a theoretical model.

The problem with an individual mind as a vector x_d is that it is all alone. An adaptive mind has to have something to adapt *to*; previous models have presumed that the mind adapts to details of the problem at hand, while the present social-psychological view presumes that minds adapt to other minds. The reasonable thing is to propose a society of minds: x_{id} minds of individuals i on dimensions d.[1] Now we have a number of individuals coexisting in a cognitive state space, each depicted as a vector of numbers, and each being an element in a vector of minds.

Change in Mental State

Development, learning, emotions, forgetting, thinking, and attitude change are some of the ways that minds change more or less constantly. The kinds of change that will be depicted in an x_{id} model depend of course on what representation was chosen (and vice versa), what dimensions are represented by the x_{id}'s.

Changes of mental state are continuous in time, in the sense that one state follows from another. Discontinuous series of mental states such as sudden disruptions of belief or mood are symptomatic of mental illness or indicative of an undetected pattern within which the new state does follow from its precedent, i.e., bad specification of the system. If the relevant variables are identified, then normal change, whatever its form, is understood to be continuous.

That is not to say that changes of mental state are predictable. The variables that affect behavior are complex, they interact in unforeseeable ways, and are introduced into the computational model as randomness. Though mental change is continuous, it is not predictable; it is continuous but random.

Continuous change in x_{id} is seen as:

$$x_{id}(t) = x_{id}(t\text{-}1) + \Delta x_{id}$$

[1] Note the contrast between a "society of mind" existing inside a skull and the present view of a society of *minds*.

leaving us with the task of explaining how Δx_{id} operates. Psychological findings suggest two major effects on Δx_{id}.

The Law of Effect

Thorndike's (1911) Law of Effect, which became the basis of subsequent reinforcement theories, stated that a response which is associated with a reward becomes more likely to recur in the future. The present sociocognitive perspective cheerfully intends to violate the behavioristic ethos by imputing mental dynamics, so in this case the "response" can be an attitude, behavior, or cognition. An individual's series of x_{id} iterated over some time-steps will have found a position p_{id} that resulted in the best evaluation so far, and toward which i will tend to return, according to the Law of Effect.

In a multivariate model such as the present one, the attraction of a previously rewarding behavior can be implemented as the vector of differences between an individual's current position and the previous best position p_{id}. Each term will be weighted by a positive random number φ, whose upper limit will be a parameter of the system, though in the following examples it will retain the limit of 2.0, so that the mean weight = 1.0. Thus, the Law of Effect is defined as an attraction by the individual toward a point p_{id}, where the attraction is proportional to $\varphi\cdot(p_{id} - x_{id})$, with a new random number φ generated each time it occurs.

Social Interaction

Sherif (1936) studied the formation of norms using the "autokinetic effect." A point of light projected on the wall of a perfectly dark room appears to move because of adaptive movements of the subject's eyes. If a subject is asked, for instance once every minute, to report how far the light has traveled, he or she will give relatively consistent answers. Subjects typically report the light moving approximately six inches per minute.

If subjects give their reports simultaneously with someone else who initially reports movements of several feet or more, however, the reported distances begin to shift, until the subjects report approximately the same amount of movement of the light as reported by the other person. Sherif's early research demonstrated the formation of norms; subsequent studies of numerous judgments and behaviors have supported the finding that individuals tend to converge in their responses to many tasks (cf. Petty and Cacioppo, 1981).

Latané and his colleagues (cf. Nowak, Szamrej, and Latané, 1990) have shown the correlation of arbitrary multiple attitudes as a function of highly persuasive individuals. The current view, however, considers attitudes and other cognitive elements to be inherently correlated, as their combination determines the goodness of a mental state. Certain patterns of thought "go together" sensibly and comfortably (as measured by $G(\mathbf{x})$), and people seek to find those patterns, though finding them is often difficult. It is suggested here that people succeed at this cognitive optimization through collaboration.

An individual communicates repeatedly with some number of selected individuals in a neighborhood, which is arbitrarily defined by position in the array of individuals. In a typical particle swarm experiment, individual i exists in a neighborhood with i-1 and i+1. Individual i will be influenced by the member of the neighborhood who has attained the best goodness measure so far: the variable g takes on the value of the index of that "best" neighbor, so that p_{gd} indicates the best position attained so far by any member of the neighborhood.

It is then a simple matter to attract i toward the neighborhood best position by a term similar to that used for the Law of Effect. The complete formula will have two terms: first, the Law of Effect term given above, and second the social influence term just described, weighted by a positive random number as before:

$$\varphi\cdot(p_{gd} - x_{id})$$

Thus, if $x_{id}(t) = x_{id}(t\text{-}1) + \Delta x_{id}$, then we theorize that Δx_{id} is a function of $(\varphi\cdot(p_{id} - x_{id}) + \varphi\cdot(p_{gd} - x_{id}))$. The order of the function is not clear however. In the following passages, the phrase "adaptation vector" will refer to:

$$\varphi\cdot(p_{id} - x_{id}) + \varphi\cdot(p_{gd} - x_{id})$$

where a new random weight φ is generated each time it occurs, that is, for each of the two terms within every *id*. This vector represents the difference between the individual's position in hyperspace and that individual's previous best position, plus the difference between the individual's current position and the neighborhood best.

In the following sections, two social adaptation models are tested and compared, using a cognition-like problem – random parallel constraint satisfaction networks. Drawing from Kauffman's (1995) NK landscapes, networks were randomly generated which contained twenty nodes, with each node having k connections from other nodes. Networks with k=2, k=4, k=10, and k=15 were tested. Connections between nodes were assigned random weights in [-1.0, +1.0], and connections were asymmetrical, that is, $w_{ij} \neq w_{ji}$. The objective function was a global one which is widely applied to Hopfield networks:

$$G = \sum_i \sum_j w_{ij} a_i a_j$$

where a_i is the activation level of node i, limited by a logistic function to [0.0, 1.0], and w_{ij} is the strength of the weight connecting node j to node i. The object is to maximize the output of this function (though the present implementation minimized its negative). Optimal vectors of nodes maximize the sum of the products of connecting nodes and their weights, so that two things which go together and have a positive weight between them should both be active simultaneously, while negatively connected nodes should turn one another off.

Increasing the numbers of connections to each node increases the complexity of the network by introducing conflicting constraints, increasing epistasis in the network. Because the networks are random, the optima are not known, of course. Each version was run twenty times, and the first column of numbers presented below indicates the mean numbers of iterations required for some member of the population to attain the best score found. The system was allowed to run for 300 iterations with $k<10$, and 500 iterations with $k\geq 10$, in order to make sure that further progress was not going to occur. The second column indicates the last iteration where any member of the population had a value less than the population best – measured in floating-point precision, this is a very strict standard. The final column gives the mean optimal or best G. As k increases, we expect G to decrease.

First-order Adaptation: Change of Position

The first experiments tested a model with the adaptation vector $\varphi\cdot(p_{id} - x_{id}) + \varphi\cdot(p_{gd} - x_{id})$ simply added to the individual's position. That is, in this model:

```
For i = 1 to number of individuals
    For d = 1 to number of weights in a network
        x_id = x_id + φ·(p_id - x_id) + φ·(p_gd - x_id)
    Next d
Next i
```

(Note: for-next loops will be assumed in succeeding examples.) Further, if a term in the adaptation vector exceeded $\pm$ 4.0, then that value was cut off to $\pm$VMAX=4.0, i.e., $x_{id} = x_{id} + VMAX_{id}$.

This model was run for 20 trials at each level of k. As seen in Table 1, a mean of 224.50 trials were required to attain the optimum when k=2, and after 239.70 iterations the entire population had attained the optimum. Interestingly, the numbers of iterations required did not increase monotonically with k, though the optimal goodness G decreased as expected.

Table 1. First-order adaptation. "First optimum" refers to the mean number of iterations required for a member of the population to attain the optimum or best score. "Last suboptimum" denotes the last iteration in which any member of the population performed at a level lower than the best. G is the mean goodness of the networks, for each value of k.

k	First optimum	Last suboptimum	G
2	224.5	239.70	0.1574
4	250.55	274.70	0.1071
10	255.3	270.00	0.0725
15	253.45	259.35	0.0516

Each individual in the first-order implementation of the adaptation vector searches within a hyperrectangle (for lack of a better term: a multidimensional volume with perpendicular corners and unequal sides) with one corner at x_{id} and the diagonal corner at:

$x_{id} + lim(\varphi)\cdot((p_{id} - x_{id}) + lim(\varphi)\cdot(p_{gd} - x_{id}))$, or

$x_{id} \pm VMAX_{id}$, whichever has a smaller absolute value.

The random numbers φ throw the particle somewhere into the box defined by those corners.

Convergence of the population was seen in these trials, as individuals gradually approached the optima, and neighbors tended to produce similar solution vectors to one another. Phenomenologically this model simulates a population of individuals who move directly to compromises between their own good solutions and their neighbors', sometimes moving to more extreme positions than the target. Upon detecting a discrepancy between their current positions and their previous bests, as well as their neighbors' previous bests, individuals change to approximate those relatively good positions in the problem space.

Second-order Adaptation: Change of Change

Next a program was written wherein individuals adapted their rate of change, as compared to their position, to the adaptation vector. That is:

$$\Delta(\Delta x_{id}) = \varphi\cdot(p_{id} - x_{id}) + \varphi\cdot(p_{gd} - x_{id})$$

or in other words:

$$x_{id}(t) = x_{id}(t-1) + \Delta x_{id}(t-1) + \varphi\cdot(p_{id} - x_{id}) + \varphi\cdot(p_{gd} - x_{id})$$

In this second-order adaptation algorithm, individuals adapt by changing the *rate* of their changing. In this version an individual's trajectory continues its course, with adjustments. Δx_{id} was limited, as before, to the range [-4.0, +4.0].

As seen in Table 2, optima which were comparable to those found in the first-order model were found in less than half the time. A multivariate analysis of variance was performed, with three dependent variables: First Optimum, Last Suboptimum, and G, and two independent variables k and Order (first- versus second-order adaptation), plus their interaction. The MANOVA (using Wilks' Lambda) was significant for Order, $F(3, 150) = 505.4720$, $p<0.0001$, and for k, $F(9, 365.21) = 18.2465$, $p<0.0001$, but not for their interaction. Univariate main effects of k were significant for First Optimum, $F(3, 152)=6.55$, $p<0.0003$, for Last Suboptimum, $F(3, 152)=6.76$, $p<0.0003$, and for G, $F(3, 152)=66.61$, $p<0.0001$. Main effects for Order were significant for First Optimum, $F(1,152)=940.87$, $p<0.0001$, and for Last Suboptimum, $F(1,152)=661.06$, $p<0.0001$, but not for G, that is, solutions were found faster and converged faster, but were not significantly better in the second-order version. No interactions were significant, suggesting that the effect of k was the same for first- and second-order algorithms.

Finally, the number of iterations between the first discovery of an optimum and the last suboptimal score differed significantly between Orders, $F(1, 152)=164.18$, $p<0.0001$, but not by k or the interaction; it took significantly longer for the second-order algorithm to converge once the optimum was found, as particles' trajectories returned toward the best positions by a more gradual route.

Table 2. Performance of second-order adaptation algorithm.

k	First optimum	Last suboptimum	G
2	104.25	128.90	0.1560
4	114.55	143.45	0.1105
10	121.45	149.95	0.0762
15	115.10	145.20	0.0614

In this second-order implementation, the particle searches a hyperrectangle with corners at:

$$x_{id}(t\text{-}1) + \Delta x_{id}(t\text{-}1)$$

and

$$x_{id}(t\text{-}1) + \Delta x_{id}(t\text{-}1) + lim(\varphi)\cdot((p_{id} - x_{id}) + lim(\varphi)\cdot(p_{gd} - x_{id}))$$

or

$$x_{id}(t\text{-}1) \pm \mathrm{VMAX}_{id}$$

whichever absolute value was smaller.

In this version the hyperrectangle may contain the particle's current position. As the particle moves ahead on the trajectory established at the previous time step before making an adjustment back toward the best points, the attraction tends to be roundabout, with particles looping past optima and returning to them.

In these models individuals are attracted toward their own previous successes, and toward the previous successes of their neighbors. The difference between the first- and second-order models lies in perseverance (Ross, Lepper, and Hubbard, 1975). In first-order adaptation, individuals abandon their previous trajectories when they or their neighbors discover a newer, better pattern of elements. In the second-order algorithm, individuals continue on their current trajectories, modifying these on the basis of new findings. The first case is as if a scientist, upon finding some new research results, gave up his or her previous goals to begin studying the new phenomenon. In second-order adaptation, the scientist continues his or her research program, but takes the new findings into account. Second-order adaptation appears to correspond well with current social psychological findings.

The Emergence of Culture

The second-order algorithm above, called particle swarm adaptation, can be implemented using any kind of cognitive theoretical model or intelligent agent architecture. For instance, it appears that performance of the particle swarm on feedforward neural nets is approximately equal to backpropagation of error, and anecdotal reports have it outperforming backprop in some cases. The method has also been shown to perform well on Hopfield networks, quantitative balance theory models, and fuzzy cognitive maps, as well as on symbolic representations such as graph-search problems.

After some number of iterations the system is seen to have converged on one or more optima. In cases where multiple global optima are discovered by the population, series of topological neighbors tend to cluster in the same optimal regions of the search space. These series extend beyond hard-coded neighborhoods. An individual which has found an optimal combination of elements draws its adjacent neighbors toward itself; if the region is superior, then the neighbors' evaluations will improve as well, and they will attract *their* neighbors, and so on. If another subset of the population is attracted to a different but equally good position, then a natural separation of groups is seen to emerge, each with its own pattern of coordinates which may easily be thought of as norms or cultures. When one solution is better than another, it usually ends up swallowing the lesser pattern, though in some cases individuals on the borders of groups (who generally perform poorly themselves) prevent the spreading of better solutions through the population.

Table 3. An example of culture in a particle swarm trial. Individuals' activation vectors (rows in the table) resemble those of their neighbors.

i	a_1	a_2	a_3	a_4	a_5
1	1.00	1.00	1.00	0.00	1.00
2	1.00	1.00	1.00	0.00	1.00
3	1.00	1.00	1.00	0.00	1.00
4	0.00	0.00	1.00	1.00	0.00
5	0.00	0.00	1.00	1.00	0.00
6	0.00	0.00	1.00	1.00	0.00
7	0.00	0.00	1.00	1.00	0.00
8	0.00	0.00	1.00	1.00	0.00

The polarization of these artificial populations into separate cultures appears very similar to the convergence of human populations on diverse norms of attitude, behavior, and cognition. Human interaction results in conformity or convergence on patterns which are similar for proximal individuals and different between groups.

The "Immergence" of Intelligence

Local interactions in complex systems result in emergent global behavior which is not predictable from the local behavior. The formation of cultures in particle swarm trials is not specified in the computer programs and is not predictable from the definitions of interactions in the programs.

Culture emerges from local interaction – but that in itself is not especially interesting for a scholar seeking to understand intelligence. A second feature of the behavior of a particle swarm system, or of a human society, is the *immergence* of cognitive adaptation as a result of the top-down effect of *emergent* culture.

The cultural convergence of individuals in the search space results in intensive exploration of optimal regions. Relatively good combinations of elements, which in human society may be beliefs, behaviors, problem-solving steps, opinions, etc., receive relatively focused attention. As a result, the performances of individuals are improved. Culture, the result of emergent bottom-up processes, is the cause of immergent mental phenomena, optimizing the cognitive processes of individuals. Culture allows intelligent behavior of individuals.

Minds and Cultures

The present paper attempts to begin to develop a theory of minds in a field containing other minds, some data, and some formula for processing data– a cognitive model, which might vary from case to case.

In this simplified model, all individuals input the same data; of course in nature each individual perceives each event uniquely, with input data determined not only by physical position in the environment but by previous learning. The treatment of environmental data is not addressed at all in the present model, though it is clear that differences between data sets can explain a lot of the observed differences between individuals.

The present implementation simulates instead the processing of a single set of data by a population of individuals. It is seen that individuals' conformity to the influence of their neighbors can result in optimal information patterns. Cultural patterns emerge from local interactions, and in turn influence the intelligent behaviors of individuals locally.

Intelligence is a label to describe the qualities of a good mind, as judged by the person using the term. No universal definition of the concept is likely to be invented, as the judgment depends on the person making it. The qualities of a mind, further, are hypothesized to immerge from the culture within which it functions. The pursuit of "intelligent" computer programs should focus on the social interactions which result in the emergence of culture and the immergence of cognitive adaptation. A versatile framework for the depiction of mental states was presented, and the optimizing effect of social interaction was demonstrated in a parallel constraint satisfaction paradigm.

A social-psychological insight is that a great amount of social interaction consists, not especially in the communication of facts, but in the communication of methods for processing the information. Through approximation of others' relatively successful cognitive methods, individuals are able to behave intelligently; the result of local interaction is culture, and the result of culture is intelligent performance by individuals.

References

Allport, G. (1985). The historical background of social psychology. In G. Lindzey and E. Aronson (Eds.), *Handbook of Social Psychology, Volume 1.* New York: Random House.

Kauffman, S. (1995). *At home in the universe: The search for the laws of self-organization and complexity.* New York: Oxford University Press.

Kelley, H. H. (1973). The process of causal attribution. *American Psychologist, 28*, 107-128.

Kennedy, J. (1996). Bad thinking, good problem-solving: human irrationality and computational intelligence. *Intelligent systems: A semiotic perspective. Proceedings of the 1996 International multidisciplinary conference. Volume II: Applied Semiotics*, 361-366.

Kuhn, T. S. (1970). *The structure of scientific revolution, 2nd ed.* Chicago: The University of Chicago Press.

Lane, H. (1976). *The Wild Boy of Aveyron.* Cambridge: Harvard University Press.

Latané, B. (1981). The psychology of social impact. *American Psychologist, 36*, 343-356.

Lillard, A. S. (1997). Other folks' theories of mind and behavior. *Psychological Science, 8*, 268-274.

Nowak, A., Szamrej, J., and Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review, 97*, 362-376.

Petty, R. E., and Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches.* Dubuque: Wm. C. Brown Company.

Ross, L., Lepper, M., and Hubbard, M. (1975). Perseverance in self-perception and social perception: biased attribtutional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*, 880-892.

Ross, L. D. 1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology, Vol 10.* New York: Academic Press.

Sherif, M. (1936). *The psychology of social norms.* New York: Harper Brothers.

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence.* New York: Penguin Books.

Thorndike, E. L. (1911). *Animal intelligence.* New York: Macmillan