An Emergentist Account of Collective Cognition in Collaborative Problem Solving

John Voiklis (jv37@columbia.edu)
Cognitive Studies in Education
Teachers College, Columbia University, New York, NY10027

Manu Kapur (manu_kapur@hotmail.com)
Communication, Computing, and Technology in Education
Teachers College, Columbia University, New York, NY10027

Charles Kinzer (kinzer@tc.columbia.edu)
Communication, Computing, and Technology in Education
Teachers College, Columbia University, New York, NY10027

John Black (black@exchange.tc.columbia.edu)
Cognitive Studies in Education
Teachers College, Columbia University, New York, NY10027

Abstract

As a first step toward an emergentist theory of collective cognition in collaborative problem solving, we present a proto-theoretical account of how one might conceive and model the intersubjective processes that organize collective cognition into one or another—convergent, divergent, or tensive—cognitive regime. To explore the sufficiency of our emergentist proposal we instantiate a minimalist model of intersubjective convergence and simulate the tuning of collective cognition using data from an empirical study of small-group, collaborative problem solving. Using the results of this empirical simulation, we test a number of preliminary hypotheses with regard to patterns of interaction, how those patterns affect a cognitive regime, and how that cognitive regime affects the efficacy of a problem-solving group.

Introduction

Collaborative problem solving presents a coordination challenge (Lewis, 1969): the timing and efficacy of top-down processes—the means-ends operations—whereby a group wends its way through the problem space, depend on the timing and efficacy of bottom-up processes, whereby heterogeneous agents evolve and propagate shared intentions, goals, beliefs, and conceptions (Clark & Brennan, 1991). Absent this intersubjective convergence (cf. Roschelle, 1992), collaborators cannot define (perhaps not even recognize) the problem at hand nor select among the possible solutions, much less take action (Katz & Lazarsfeld, 1955). Luckily, collaborators tend, over time, toward psychological homogeneity (for review see Arrow, McGrath, & Berdal, 2000). In fact, human beings appear hard-wired for sharing psychological states (Tomasello et al., 2005); the urge to converge has adaptive value—i.e. survival of the groupiest (e.g. Axelrod, 1984). That said, the efficiency gains afforded by complete and uncritical consensus—groupthink—come with adaptability costs, as well (Janis, 1982); collaborators proceed with cognitive myopia, leaving much of the problem and solution spaces unexplored. Some groups manage to generate and sustain sufficient tension between the intersubjective convergence necessary for concerted action and the divergence necessary for cognitive flexibility; other groups lurch towards one extreme or another. How do groups tune the levels of their collective cognition; why and when do some groups succeed and others fail?

Collective Cognition: Supervenient, Yet Distinct

One word, emergence, provides an easy answer. Collective cognition—whether convergent, divergent, or at some tension point between the two—emerges from the intersubjective interactions—discussion, negotiation, and speculation—among collaborating agents; each interaction, in relation to every other interaction, tunes collective cognition both in time and over time. Invoking emergence only begs the question: how, when, and why does collective cognition emerge? Among cognitive scientists, interest in collective cognition and its emergence is a recent phenomenon (e.g. Goldstone, 2005); existing theories (e.g. Hutchins, 1995) detail how collective cognition propagates once structured and institutionalized, but a theory of its emergence remains forthcoming (Schwartz, 1995). The need for a theory stems from the cumulative effect of empirical research indicating that intersubjective processes yield cognitions—e.g opinions (Isenberg, 1986), knowledge representations (Schwartz, 1995), decisions (Bornstein & Yaniv, 1998), among others—that differ, both in complexity and kind, from those produced by any collaborating agent or those expected from the central tendency among collaborators (Vallacher & Nowak, in press). Moreover, these group-level cognitions emerge spontaneously, without forethought or awareness among collaborating agents (Goldstone, 2005). Apparently, the interactions among collaborators generate a mind—supervenient, yet distinct from any constituent mind—requiring cognitive inquiry in its own right.
Emergence: Unpacking the Paradox

Despite newfound enthusiasm, the road to an emergentist theory of collective cognition remains littered with obstacles. Foremost, emergence presents a definitional paradox: on one level, emergent phenomena depend on underlying variables and processes; on another level, emergent phenomena remain autonomous from underlying variables and processes (Bedau, 2003). This paradox complicates the ontological, causal, and explanatory derivation of emergent phenomena. First, the ontological derivation of an emergent phenomenon is informationally complex (ibid.). This means that one could reduce collective cognition to its constituent psychological variables (as manifest by intersubjective interactions), however, one could not derive collective cognition without full information on how those variables aggregate and interact. Second, the causal derivation of an emergent phenomenon is informationally complex (ibid.). One could reduce collective cognition to its current configuration of psychological variables, but one could not derive collective cognition without full information on the history of intersubjective interactions that led to that configuration. Third, no particular ontological or causal derivation can explain an emergent phenomenon (ibid.). Over the duration of the problem-solving process, collective cognition may organize into a variant of three—convergent, divergent, or tense—attractors or change-resistant cognitive regimes. Apparently similar configurational histories may organize very different cognitive regimes, while apparently different configurational histories may organize similar cognitive regimes. Fourth, an emergent phenomenon propagates the ontological and causal conditions from which it derives (ibid.). Intersubjective interactions may organize collective cognition into a particular regime, but the regime downwardly constrains the intersubjective interactions by which collective cognition propagates. All in all, an emergent phenomenon, itself, offers the shortest description—ontological, causal, and explanatory—of its own emergence (ibid.); collective cognition is algorithmically irreducible.

Deriving the Irreducible: A Minimalist Approach

The algorithmic irreducibility of emergent phenomena precludes any short cut derivation of collective cognition: one must recapitulate the full ontological and causal history from which collective cognition emerged—i.e. derivation by simulation (ibid.). At first glance, derivation by simulation appears tedious, perhaps intractable: for one, the full ontological and causal history of collective cognition involves multi-level causal dynamics operating on multiple, permuting variables; further, as argued above, no particular ontological or causal derivation can explain collective cognition. Algorithmic irreducibility, one might suspect, also precludes a parsimonious theory of collective cognition. Then again, the ontological and causal history of collective cognition entails, at each point in time, a finite set of local intersubjective interactions; each local interaction entails only local dynamics operating on local variables. This enables a minimalist approach to derivation by simulation (cf. Nowak, 2004): one need only model the minimal information—e.g. utility function, decision rule, or heuristic—contained in a local interaction; through repeated updating of aggregated local interactions, the simulation generates the phenomenon—in all its informational complexity—from the bottom up (ibid.). Employing a minimalist, bottom-up approach to derivation by simulation, an emergentist theory of collective cognition need not sacrifice explanatory depth and richness for parsimony (ibid.); Ockam can put away his razor.

Computational Models: Life-Like Is Like Life

Computational models of collective behavior and collective psychology exemplify the minimalist, bottom-up approach to derivation by simulation. (for reviews see Goldstone, 2005 and Vallacher & Nowak, in press). For instance, the dynamical implementation of Social Impact Theory (Nowak, Szamrej, & Latané, 1990) simulates how polarized clusters emerge in public opinion. In the theory and in the simulation, social influence operates via two interlocking mechanisms: the group influences each person, and each person influences the group. The intensity of that influence, both group-on-person and person-on-group, derives from a function of three variables: group size, personal persuasiveness, and personal position in physical (or social) space. During the course of discussion—i.e. the iterative application of the social influence function to each group-on-person and person-on-group interaction—the simulation evolves from an initial random distribution of opinions to a distribution of opinions not unlike that in the real world: islands of minority opinion in a sea of majority opinion. One finds similarly plausible patterns of collective behavior and thought in simulations that model higher-dimensional cognitive structures, yet lower-dimensional mechanisms of social influence; global systems of cultural knowledge (Kennedy’s, 1998) and meaning (Barr, 2004) can emerge—with surprising efficiency—from local, person-to-person exchanges of partial knowledge. Verisimilitude—the plausibility of behavior and thought patterns—lends explanatory power to computational models. If simple mechanisms operating on minimal variables produce realistic phenomena in a simulated world, perhaps the same simple mechanisms operating on the same minimal variables produce real phenomena in the real world (Nowak, 2004).

Two Routes to Theory Building

Computational models, thus, offer one way to understand how, when, and why a collaborative, problem-solving group succeeds or fails in tuning the levels of its collective cognition. One could implement any number of minimalist, hypothetical models and validate the results against empirical data (Goldstone, 2005). The model whose aggregate iteration produces the most plausible patterns of collective behavior and thought likely underlies the process by which real-world collective cognition emerges and propagates (Nowak, 2004). Thus, verisimilitude proves essential to the theory-building efforts of those trying to understand collective behavior and collective psychology.
That said, a reliance on verisimilitude may strain one’s evidentiary standards (Goldstone, 2005); for instance, while one might rely on the verisimilitude of Shakespeare’s *King Lear* to make sense of observed familial dysfunction, one might hesitate to claim that the contrived vicissitudes of a fictional family—no matter how plausible—can explain the observed vicissitudes of an actual family. In response to this wariness, one could follow an empirical route to derivation by simulation: beginning, again, with a minimalist, hypothetical model of a local intersubjective interaction, one could use empirical (as opposed to computer-generated) data to update the aggregated local interactions at each iteration of the simulation. The empirical route, though, comes with its own limitations. For one, life rarely maintains a time-ordered log of intersubjective interactions; one must rely on data from laboratory experiments, where sample size, time scale, and transparency rarely match the levels available in computational experiments. Further, the intersubjective interactions of human collaborators involve a panoply of discursive instruments—analyses, jokes, lies, as well as propositions; before proceeding with simulation, one must translate these discursive instruments—via some theory-based process—into simple mechanisms operating on minimal variables. Despite these limitations, empirical simulations avoid validation by verisimilitude; the simulated patterns of behavior and thought are not simply life-like, they are life. One can expect empirical validation to supplement, if not supplant, verisimilitude in building an emergentist theory of collective cognition in collaborative problem solving.

**Collective Cognition: An Emergentist Proposal**

Before one can proceed—whether via computational or empirical simulation—with building a emergentist theory of collective cognition in collaborative problem solving, one needs a proto-theoretical account of the phenomenon to guide how one might conceive and model the intersubjective interactions whose aggregate iteration organize collective cognition into one or another—convergent, divergent, or tensive—cognitive regime. While speculative, such an account would allow one to generate a number of preliminary, yet testable, hypotheses with regard to patterns of intersubjective interaction, the effect of those patterns on the stability of a cognitive regime, and the effect of a cognitive regime on the problem-solving efficacy of a collaborative group. To that end, we offer the following emergentist proposal:

How, when, and why does a collaborative, problem-solving group succeed or fail in tuning the levels of its collective cognition and, thereby, the levels of its problem-solving efficacy? As indicated earlier, a collaborative, problem-solving group must coordinate the bottom-up manipulation of intersubjective variables—intentions, goals, beliefs, and conceptions—with the top-down manipulation of problem-related or instrumental variables. Both manipulations operate via the same mechanism, intersubjective interaction—the discussion, negotiation, and speculation through which collaborating agents generate and enact a shared representation of the problem and its solution. Each successive interaction reconfigures both intersubjective and instrumental variables in such a way that may, in relation to the reconfigurations of previous interactions, increase or decrease the need for further discussion, negotiation, and speculation: i.e. each successive interaction impacts the level of convergence among collaborating agents and, thereby, the efficiency of the ensuing problem-solving process. Given the coordinative interdependence between instrumental and intersubjective levels of convergence, one can expect interactions that generate convergence/divergence on one level to follow on interactions that generate convergence/divergence on the other level. Thus, each interaction may constrain the valence—convergent, divergent, or neutral—of the succeeding interaction, creating clusters of interactions with a similar impact on both convergence levels and efficiency. These clusters or tiny attractors (cf. Kauffman, 1993), in turn, constrain the overall trajectory of the problem-solving process: recurring tiny attractors push the discussion, more and more, in one direction, organizing a cognitive regime—i.e. a major attractor. With the regime in place, similarly valenced interactions—both singly and in clusters—increase in likelihood; while one might observe convergent interactions in a divergent regime or divergent interactions in a convergent regime, these perturbations have little impact on the direction of the discussion. A group whose cognitive regime tends towards divergence remains mired in disagreement and indecision: collaborating agents cannot find a mutually satisfactory solution. A mutually satisfactory solution comes easily for collaborating agents whose cognitive regime tends towards convergence, but that solution reflects the monolithic tendencies of the group. A multidimensional solution requires multidimensional intersubjective processes; one can expect such a solution from a group whose collective cognition tends toward a tensive cognitive regime.

**Purpose**

In what follows, we explore, via empirical simulation, the sufficiency of our emergentist proposal of how, when, and why a collaborative, problem-solving group succeeds or fails in tuning the levels of its collective cognition and, thereby, the levels of its problem-solving efficacy. Specifically, we instantiate one minimalist model of intersubjective convergence and simulate the tuning of collective cognition using data from an empirical study of small-group, collaborative problem solving. Finally, we present a sequence of explorations, each of which tests, using the results of the empirical simulation, one or two preliminary hypotheses.

**Method**

**Research Context and Data Collection**

The data for this empirical simulation come from a study of collaborative triads solving problems in an online, synchronous chat environment. That study looked at the effects of problem structure—e.g. well-structured or ill-structured problems—on the nature and efficacy of
computer-mediated collaboration; problem structure functions as a control variable in our explorations. Participants included sixty 11th grade students (46 male, 14 female; 16-17 years old) from the science stream of a co-educational, English-medium secondary school in Ghaziabad, India. They were randomized into twenty triads, each of which had to collaborate in solving a well-structured or ill-structured problem scenario. Both problems asked groups to determine liability in an automobile accident; solutions required the application of Newtonian kinematics. The study took place in the school’s computer laboratory, where collaborators communicated with one another entirely through synchronous, text-only chat. The chat server archived a time-ordered transcript of each group’s discussion. These twenty transcripts serve as the data driving both the simulation and subsequent analyses.

**Operationalizing Intersubjective Convergence**

As a conceptual domain, Newtonian kinematics is ontologically direct: known problems have a small set of known solutions. In solving kinematics problems, a group’s intersubjective processes must converge on this small set of normative concepts, strategies, and solutions: i.e. the instrumental and intersubjective dimensions of the problem-solving processes collapse to one dimension. Consequently, one can think of the problem-solving process as a walk along a straight path: success and failure await at either end. With each intersubjective interaction, a group may step forward (convergent valence), step backward (divergent valence), or stand still (neutral valence). The mean distance traveled along this path represents a group’s overall convergence—their proximity to the small set of normative concepts, strategies, and solutions. Operationally, then, one can model both the problem-solving process and the resulting overall convergence as a Markov walk (Ross, 1996).

For the purposes of empirical simulation, two trained doctoral students independently segmented the twenty transcripts into semantically-defined, interaction units: i.e. each utterance was divided into each constituent phrase—variable identification, strategy suggestion, solution evaluation, et al.—that could impact the group’s level of convergence and problem-solving efficiency. Following segmentation, the coders assigned an impact value of 1, -1, or 0 to each interaction unit (Krippendorff’s alpha = .93) depending upon whether, in relation to previous interactions, the interaction represented a step forward (impact = 1), a step backward (impact = -1), or no step at all (impact = 0). In this way, each discussion, with its panoply of discursive instruments, was reduced to a temporal string of 1s, -1s, and 0s. At any point in the problem-solving process, a group’s overall convergence derives from a function of those of 1s, -1s, and 0s. At any point in the problem-solving process, a group’s overall convergence derives from a function of those of 1s, -1s, and 0s. More formally, at any point in the problem-solving process, let $n_1$, $n_2$, and $n_0$ denote the number of interaction units assigned an impact value of 1, -1, and 0, respectively. Then, up to that point, the convergence value would equal, $C = (n_1 - n_2)/(n_1 + n_2)$. To run the simulation, we calculated a convergence value after each utterance; for each of the 20 discussions, this generated a notional time series representing the evolution of each group’s collective cognition.

**Operationalizing Problem-Solving Efficacy**

Throughout our explorations, a group’s ultimate, problem-solving efficacy—the accuracy and quality of its solution—serves as either a criterion or predictor. As argued earlier, multidimensional solutions indicate an efficacious problem-solving process: i.e. an efficacious group will produce not only a correct solution, but a solution with a variety of quantitative and qualitative arguments. Two doctoral students independently assessed the accuracy and quality of each solution based on a nine-point rubric (Krippendorff’s alpha = .97).

**Results and Discussion**

**Exploration 1: Interpreting the Fitness Curve**

![Figure 1: Four illustrative fitness curves.](image)

Plotting the convergence value on the vertical axis and time (defined notionally, with each utterance a tick on an evolutionary clock) on the horizontal axis, one can visualize the evolution of each group’s collective cognition (Figure 1). Three aspects—end point, length, and shape—of this visualization or fitness curve appear informative. The end point of each curve indicates the final level of convergence, from which one can deduce each group’s proximity to the small set of normative concepts, strategies, and solutions. The length of each curve indicates the duration of each problem-solving process: i.e. the efficiency with which each group reached its final solution, however proximal to the normative set.

The shape of each fitness curve appears the most informative aspect of all. For instance, the early portion of each curve indicates whether apparently similar initial processes lead to different trajectories, while apparently different initial processes lead to similar trajectories. In Figure 1, both of the low-efficacy groups—whether solving a well- or ill-structured problem—tuned their collective cognition to similarly low levels of convergence; yet, each arrived at those final levels via different paths. After some initial positive steps, the convergence level among collaborators in the Low/Well group declined sharply then plateaued. The Low/Ill group, on the other hand, appeared to recover from their initial missteps, before they too plateaued in divergent terrain. These fitness plateaus—evident both among the high-efficacy and low efficacy groups—offer
preliminary evidence for the existence of cognitive regime and its effects on a group’s problem-solving efficacy: a convergent regime among groups with high-efficacy outcomes, a divergent regime among groups with low-efficacy outcomes. One should note that no group plateaued at one extreme or another: as one would expect from interacting human beings, low-efficacy groups generated some minimal level of convergence; more importantly, high-efficacy groups appear to have sustained enough divergence to explore a wide swath of the problem and solution spaces.

Given this visual evidence for an early-emerging cognitive regime and its correspondingly early effects on problem-solving efficacy, we now explore this relationship through statistical means.

**Exploration 2: The Cognitive Regime**

One way to verify the existence and effect of a cognitive regime involves testing how early and how consistently one could predict (with p-values ≤ 0.05) a group’s eventual problem-solving efficacy from the level of convergence. For the purposes of this test, we segmented the twenty discussions into ten equal parts; then, at each tenth, we calculated the convergence value up to that point. This resulted in ten sets of twenty convergence values: the first set (C1) corresponding to convergence after 10% of the discussion, the second set (C2) after 20% of the discussion, and so on until the tenth set (C10), corresponding to the final convergence value. To simulate the temporal effects of cognitive regime we regressed problem-solving efficacy on C1, then C2, and so on through C10, controlling for problem structure each time.

Figure 2: Simulating the predictive power of convergence.

To visualize the simulation we plotted the *p*-value that corresponded to the statistical significance of each regression (Figure 2). The data suggest that, beginning at some point between 30% and 40% into the problem-solving process (between C3 and C4) through to its end, the level of convergence can, on average, predict (with p-values ≤ 0.05) the eventual problem-solving efficacy. The simulation confirms not only that interacting agents do, in fact, organize their collective cognition into a cognitive regime, but that they do so early and with consistent consequences.

Having verified the early emergence and predictive power of a cognitive regime or main attractor, we next identified and assessed the tiny attractors that organize the regime and maintain its immunity to perturbations.

**Exploration 3: Tiny Attractors**

Earlier, we defined tiny attractors as sustained sequences or clusters of interactions with the same valence. Lag Sequential Analysis (LSA) detects various non-random patterns in a given sequence of interactions; specifically, we looked for statistically significant autocorrelations (Bakeman & Gottman, 1986)—instances where an interaction with a particular impact value followed on an interaction with the same impact value. From these statistically significant autocorrelations, one might induce that similarly-valenced interactions were sustained, with greater likelihood, in stochastic clusters (tiny attractors) rather than spread randomly throughout the discussion. As expected, high-efficacy groups were 133% more likely to sustain convergent interactions, while low-efficacy groups were 120% more likely to sustain divergent interactions. In each case, recurring tiny attractors pulled the discussion into one or another cognitive regime with the expected efficacy outcomes.

At this point, one might wonder whether this emergentist account of collective cognition and its derivation by empirical simulation add anything, over and above previous research, to one’s understanding of collaborative problem solving? To answer this question, one would need to instantiate and test the minimalist model of intersubjective convergence in a number of other problem-solving contexts. That said, we provide a preliminary answer by comparing the predictive power of our convergence model with that of other commonly used predictors.

**Exploration 4: Convergence vs. Other Predictors**

Previous research suggests several ways to model the problem-solving process as a function of convergent and divergent interactions. These models rarely account for the full ontological and causal history from which collective cognition emerged. One model might account for the number of convergent interactions (*Frequency* = \( n_i \)). Another model might account for the relative number of convergent interactions (*Relative Frequency* = \( n_i / (n_i + n_o + n_d) \)), i.e. convergent interactions as a proportion of all interactions. Yet another model might account for the difference between the number of convergent and divergent interactions (*Position* = \( n_i - n_d \)). Using multiple regression, we simultaneously compared the significance of all four models—frequency, relative frequency, position, and convergence—in predicting problem-solving efficacy (controlling for problem structure).

Table 1: Regression Parameter Estimates.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>-3.000</td>
<td>1.382</td>
<td>4.716</td>
<td>0.048</td>
</tr>
<tr>
<td>P. Structure</td>
<td>0.213</td>
<td>0.236</td>
<td>0.811</td>
<td>0.383</td>
</tr>
<tr>
<td>Convergence</td>
<td>7.578</td>
<td>1.891</td>
<td>16.059</td>
<td>0.001</td>
</tr>
<tr>
<td>Frequency</td>
<td>0.019</td>
<td>0.018</td>
<td>1.109</td>
<td>0.310</td>
</tr>
<tr>
<td>Position</td>
<td>-0.040</td>
<td>0.022</td>
<td>2.505</td>
<td>0.136</td>
</tr>
<tr>
<td>Rel. Freq.</td>
<td>1.885</td>
<td>2.198</td>
<td>0.735</td>
<td>0.406</td>
</tr>
</tbody>
</table>

862
Among the four predictors, only convergence recapitulates enough of the ontological and causal history to significantly predict problem-solving efficacy ($F = 16.059, p = 0.001$): the data support our earlier contention that collective cognition is algorithmically irreducible. This result hints at the further insights to be gained from investing effort and resources in an emergentist theory and methodology.

**Implications**

Our explorations reveal a number of preliminary, yet compelling, insights into the nature and dynamics of collective cognition in collaborative problem solving. As proposed, collective cognition emerges from the intersubjective interactions among collaborating agents. Each interaction both tunes the level of intersubjective convergence and constrains each subsequent interaction: e.g. convergent interactions promote convergent interactions. These tiny attractors organize a major attractor: a cognitive regime that constrains all subsequent interactions and, thereby, the outcome of collaborative efforts. This self-organizing process takes hold early in the collaboration. Consequently, one can predict, early on, whether a problem-solving group will succeed or fail. These insights have epistemological, methodological, and practical implications. At the epistemological level, our findings challenge an epiphenomenal view of collective cognition: while supervenient on the intersubjective interaction among collaborating agents, collective cognition has very real consequences. Concomitantly, our findings support a shift an epiphenomenal view of collective cognition: whether a problem-solving group will succeed or fail. These insights have epistemological, methodological, and practical implications. At the epistemological level, our findings challenge an epiphenomenal view of collective cognition: while supervenient on the intersubjective interaction among collaborating agents, collective cognition has very real consequences. Concomitantly, our findings support a shift in thinking about the nature and dynamics of collective cognition in collaborative problem solving.

**References**


