Soccer Science and the Bayes Community: Exploring the Cognitive Implications of Modern Scientific Communication

Jeff Shrager,a Dorrit Billman,b Gregorio Convertino,b J. P. Massar,c Peter Pirollib

aStanford University Symbolic Systems Program (consulting) and CollabRx, Inc.
bPARC (the Palo Alto Research Center), Inc.
cBerkeley, CA

Received 10 July 2008; received in revised form 6 June 2009; accepted 15 June 2009

Abstract

Science is a form of distributed analysis involving both individual work that produces new knowledge and collaborative work to exchange information with the larger community. There are many particular ways in which individual and community can interact in science, and it is difficult to assess how efficient these are, and what the best way might be to support them. This paper reports on a series of experiments in this area and a prototype implementation using a research platform called CACHE. CACHE both supports experimentation with different structures of interaction between individual and community cognition and serves as a prototype for computational support for those structures. We particularly focus on CACHE-BC, the Bayes community version of CACHE, within which the community can break up analytical tasks into “mind-sized” units and use provenance tracking to keep track of the relationship between these units.

Keywords: Scientific reasoning; Distributed cognition; Provenance tracking; Decision analysis; Bayes community; Collaboration; Bayesian reasoning

The process of scientific discovery—how we do science—will change more over the next 20 years than in the past 300 years.
—Michael Nielsen (http://michaelnielsen.org/blog/?p=448)
1. Introduction

As a form of distributed analysis, science involves both individual work that produces new knowledge and collaborative work to exchange information with the larger community. There are many particular ways in which individual and community can interact in science, and it is difficult to assess how efficient these are and what the best way might be to support interaction. Toward filling these gaps in our knowledge of how science works, we have developed a research platform, called CACHE, that serves two functions. First, it supports experimentation with different structures of interaction between individual and community cognition in science. Second, it enables us to prototype computational support for those structures. This paper reports on a series of such investigations.

Our work is guided by characteristics of cognition that have substantial impact on the conduct of science. Being a distributed process that takes place at many levels, there are cognitive principles that apply at the level of working group and of community, as well as the level of the individual. Our approach in this paper is to identify those principles that operate strongly at the level of the individual, yet propagate through and affect distributed cognition. We have identified five such principles. Our goal is not to provide a complete characterization of the factors affecting scientific cognition, but to motivate and position our contribution.

1.1. ‘‘Mind-sized’’ analytical sets

Effective attentive reasoning or problem solving requires operating on small, ‘‘mind-sized’’ collections of information, or chunks. When very large information collections are potentially relevant, as is the case for science, the ability to effectively bound and select small relevant sets is very important (e.g., relevant studies from long lists of related papers from a search engine). The idea of a limited focal set of information has pervaded the contemporary understanding of cognition, going back at least as far as Miller’s classic paper (Miller, 1956).

1.2. Apt representation

An apt external problem representation both distinguishes among importantly different types of information and organizes the information in a way that both matches the structure of information and supports the operations to be conducted on it. As shown by researchers and designers in information visualization, effective representations can enable tasks that are not otherwise feasible (Card, Mackinlay, & Shneiderman, 1999). They can improve performance whether successfully carrying out calculations, or debiasing reasoning (Gigerenzer & Hoffrage, 1995; Kotovsky, Hayes, & Simon, 1985; Zhang, 1997). The importance of external representation for science, at the individual and cultural level, is widely noted (Bridewell, Sanchez, Langley, & Billman, 2006; Nersessian, 1999; Novick & Hurley, 2001; Stigler, 1984).
1.3. Efficient and controlled sharing

An individual working within a community is vulnerable to conflicting pressures between what facilitates working alone versus what facilitates interacting with the community. Analogous tensions at the individual versus group level, and attendant process costs of group work, are widely noted (Nunamaker, Dennis, Valacich, Vogel, & George, 1991). The impact of changing distribution methods is addressed in studies of the history of science and in recent work on distributed groups of scientists and engineers (Olson, Finholt, & Teasley, 2000).

1.4. Task orientation

An individual scientist not only needs to interact with a large and heterogeneous collection of information available in the community, but she or he does so for a variety of tasks, projects, and purposes. Preserving context so that information needed for the same task is organized together reduces costs of switching among activities. The value of organizing information and documents by task has been recognized for individual work (Henderson & Card, 1986). Further, the need for per-task organization is likely to increase as a person’s work becomes more linked to varied sources in a community, and as individuals are driven to impose task structure on communication tools (Bellotti, Ducheneaut, Howard, Smith, & Grinter, 2005).

1.5. Multistage inference

Complex reasoning often entails a cascade of sequential reasoning steps. Such sequences risk compounding error, either by amplifying biases in a single step or by introducing new biases across the steps. In multistage probabilistic inference, people are inaccurate primarily because they fail to aggregate the information across steps in accord with normative models (Gettys, Kelly, & Peterson, 1973; Johnson & Halpin, 1974). Johnson and Halpin (1974) call for systems that address this problem and support analysts’ multistage reasoning by decomposing complex problems into manageable chunks.

An individual scientist, endowed with these and other cognitive characteristics, carries out her (or his) work within the social structure organizing the science of her time. This social structure changes. In the next section we describe Open Science, a recent sea change in how scientific information is exchanged (e.g., Bradley, Owens, & Williams, 2008). Open Science poses both opportunities and demands. After discussing this change, we describe the CACHE series of test-beds for exploring how cognitively based technology can support a cluster of scientific activities within the emerging social form of Open Science. CACHE is a Web-based platform that supports distributed, collaborative analysis. We have used CACHE both as a platform for experimentation in Open Science and as prototype of tools that could help communities of scientists cope with the flood of information posed by the Open Science movement. Our focus in this paper is not on the CACHE tool per se, but on how social organization and computational tools can help in reducing the cognitive load imposed upon analysts by Open Science.
2. Open Science

A pervasive historical fiction uniquely honors individual intellectual contributions to science; witness prizes such as the Nobel, almost always awarded to individuals. But modern science is much too complex for any individual to make significant progress without participating in numerous, often large, collaborations. Indeed, modern science is a massive, evidence-based, nonmonotonic, multistage reasoning system, distributed across millions of scientists, thousands of institutions, and hundreds of years. What all of these individuals and groups have in common is that their progress depends upon being able to efficiently draw upon one another’s results and methods, and then to efficiently feed back high-quality knowledge into their community.

Two key classes of factors control scientific productivity: the cost of doing high-quality research by an individual or laboratory (e.g., time, instruments), and the cost of scientific communication and integration of new knowledge. The second class of factors include (a) the time and tools required to share or distribute contributions, (b) the quality of the new scientific contributions, and (c) the difficulty of finding and integrating one’s own contributions with related contributions made by others. Communication and integration of knowledge are critical because if low-quality research, regardless of its cost, floods the scientific knowledge base, time and money will be wasted in either following up false leads or in trying to sift the wheat from the chaff.

Historically, the functions of communicating and archiving knowledge among scientists have been addressed through increasingly improved tools and practices: from technical correspondence before the 1700s, to archival journals introduced in 1700s, to archival journals with assessment of originality and soundness of the contributions in the 1800s, to regular conferences in the 1900s (and preprinted conference proceedings in the late 1900s). Finally, in the 2000s, the rapid diffusion of Web technology has permitted much more efficient archiving and communication in digital libraries and databases. While the functions of communication and archiving have progressively improved, the amount of information that scientists can access has also increased. More importantly, submitting contributions has become easier and faster, leading to a larger volume of contributions and greater variability in the quality of the contributions. This places greater responsibility for assessing the quality of the contributions and checking the sources on the reader. Also more workload and tensions are placed on the review and publications process (Grudin, 2004).

Since at least the 18th century, a robust quality control system based upon peer review has protected the scientific knowledge-production system from significant waste. However, in the past decade, Web-based publications such as blogs, forums, wikis, unreviewed online databases, prepublication manuscript repositories, formal online prepublication, and unreviewed (or reader-reviewed) publications, have created a so-called Open Science. Open Science is characterized by always-on, uncensored, instantaneous, free, and ubiquitous communication. This trend appears to be an inevitable adaptation of science to the rapid spread of new technologies and emerging practices. For example, Grudin (2004) reconstructs such adaptation for Computer Science and Human-Computer Interaction, a community that was exposed to the Web technology before others.
Open Science introduces both opportunities and risks. It clearly allows for increased efficiency, because contributions can be shared and accessed faster, and at lower cost. On the other hand, if rapidly reported but low-quality publications flood the (Web-based) literature, it might lead science into a crisis of wasted effort. If such a crisis is to be avoided, social and technical quality-control structures analogous to peer review must arise. What sorts of technologies might help maintain robust scientific progress in light of changing communication?

3. Soccer Science: The cognitive implications of Open Science

As science often progresses by appeal to metaphor, we would like to contribute one of our own: Contrast American football with soccer. Although both football and soccer involve moving a ball down the field toward a goal, the manner of play is completely different. Football has a clear plan-execute cycle whose tempo enables each play in the game to be tightly controlled by the coach. The coach has a view of the whole state of play, and when a play ends, he or she interacts with the quarterback to plan the next play. The team then tries to execute the plan, engaging as needed in localized opportunistic execution. In rather short course, everything stops, the coach takes stock, plans the next play, and so on. Soccer is, by comparison, a more fluid and continuous process, which, like Open Science, can be said to be ‘‘always-on’’: Planning and execution are not sequential but parallel and interleaved with players setting up for the next play even while the current one is underway. Even though the high-level strategy and roles are set by the coach in soccer, the planning and execution of the action is in done dynamically by players on the field during the running play, and particularly by whoever has the ball. Soccer does not, indeed it cannot, have a centrally controlled, plan-execute cycle like football because there are no breaks in play when an external coach could intervene to revise the plan.

These games’ very different tempos lead to very different cognitive requirements for the players: Because any member of a soccer team might become master of the play, and must take at that moment both the strategy and tactics into his or her own hands (or, rather, his or her feet), every player must be aware of the entire state of play and tactics of the moment and must compute updated tactics at a moment’s notice. Thus, every player must understand the strategy and tactics of the game in great depth. Moreover, the lack of planning breaks in soccer deprives the players both of the leisure of having a plan before each execution, and of an overview of the state of play. In football, on the other hand, the fielded players (aside from the quarterback) operate almost entirely tactically, with strategy being mainly in the hands of the omniscient coach (and to a much lesser degree, the quarterback). A football player (aside from the quarterback) will sometimes take it upon himself to change the plan in the middle of a play, but when this happens it is considered a special case. In sum, the per-player temporal constraints, informational limitations, and knowledge requirements in soccer are enormously larger than in football. All players presumably must ‘‘satisfice’’—that is, make decisions under constraints of imperfect information and limited time (Simon, 1956), meaning that all players must understand the strategy and tactics of
soccer to great depth. As these constraints are much more severe for soccer players, we would expect that the planning and execution strategies they employ to be quite different than those of football players.

Metaphorically speaking, the always-on, uncensored, instantaneous, free, and ubiquitous nature of Open Science is changing science from a football to a soccer model. In “Football Science,” small teams plan and experiment more or less independently of what the rest of the field is doing, publishing results through the long, narrow, quality-controlled publication cycle, and only having to reconsider their work, generally, “at the breaks in play”—that is, when the journals come out. The publication cycle dominates football science and enables the independent labs to engage in a somewhat leisurely read-plan-experiment-publish cycle. In the new world of Open (“Soccer”) Science, communication is continuous and interleaved with scientific work; instead of running many experiments, and then pushing some results through a structured, peer-reviewed publishing cycle, soccer scientists could potentially publish their uncensored lab notebooks, or even raw data, directly on the Web, making every datum and interpretive thought available to every other team, sans review. From the standpoint of the receiver of the information everything must be processed as soon as possible, and on an unpredictable cycle. In the limit, the ball is being passed continuously, requiring every lab in the world to reevaluate its work on a nearly daily basis. Just as soccer places a much greater cognitive load on each individual player than does football, and presumably leads to different strategies that involve greater satisficing, one would expect Soccer Scientists to experience a much greater planning load, and that they would employ strategies that satisfice much more extensively.

While high pressure, fast action is fun in sports, it could be deleterious to science. The burden imposed by the flood of scientific information of uncertain quality, while having no more hours in a day to assess it, would tend to increase satisficing and decrease the quality of individual decisions. On the other hand, because decisions are being made much more rapidly, perhaps this decrease in the quality of individual decisions will be more than balanced by increase in the rate at which the space of solutions is searched, dropping out bad solutions rapidly, and moving rapidly to new parts of the search space. This could increase the overall quality of decisions when the community of science is taken as a whole.

Our purpose here is not to take a stand on the pros or cons of Soccer Science; we take it as a matter of fact that modern science is changing in this way. We are interested specifically in how new forms of distributed cognition evolving under these conditions may best be supported. We see several distinct but interlocked implications of this shift, from the point of view of individual or small lab-group. First, as Soccer Science will make available a much larger flow of potentially relevant information, scientists will need to maintain awareness of a larger body of information. Second, the characteristics of the information stream will change in quality and stability. Specifically, we could expect that the average quality will drop and high-quality information will be less clearly marked, as the peer-review process (or whatever form assessment of contributions takes) becomes less structured and more separated from the production of results. In addition, the information will be less stable, as quickly posted, preliminary information is more likely to be revised. These changes mean that identifying the high-quality information relevant to a particular
researcher will be more difficult and so, other things being equal, will take more effort to find. In turn, scientists will either need to spend more time finding and filtering information and less time acting on it, or act based on less complete or reliable information. As the pace of information distribution in the community speeds up, this will push an individual researcher to act faster, for example, to avoid being scooped, which may in turn reduce an individual scientist’s ability or willingness to review, reflect, or check, at the same time that external vetting is reduced.

The following vignette illustrates some of the shifting in demands imposed by Open Science:

A molecular biologist, let’s call her Anne, is trying to determine the function of a protein whose gene, Ga, has just been sequenced. (This is called “gene annotation,” and genes are said to “code for” proteins.) Genes of similar sequence are often found to have similar function, so Anne enters the sequence into a search engine. Suppose that Ga has a sequence that is similar to another gene, Gb, that was previously annotated as “glucokinase” by another biologist, Barry, and posted on his public blog-like lab notebook. Anne proceeds to post on her public blog that Ga codes for glucokinase based on the results of its match with Barry’s Gb. Other scientists read Anne’s blog and draw further conclusions based upon her assertion that Ga codes for glucokinase. Suppose that some time later a third scientist, Chris, reanalyzes Barry’s data and determines that Gb does not, in fact, code for glucokinase but rather for glyceraldehyde dehydrogenase. Chris posts this to his blog, and maybe even drops Barry a courtesy email (or, more likely, posts a comment on Barry’s blog). Now, note that Anne, and others relying on her work, cannot detect this change in reliability of Barry’s evidence. She has, at least in the current world, no way of knowing of Chris’s update to Barry’s results. Further, even if Barry changes the information he posted, there may be no reason for Anne to search again and detect the change.

Although this vignette is fictional, it is not far from reality. Although most biologists today rely upon government-maintained databases for such information, not one another’s blogs, many biologists, including one of the present authors (JS) has been tripped up by just such an updating problem in these very government-maintained databases! In the Open Science world, when the interaction is through blogs and search engines, this would be a far more pervasive problem.

In short, the shift to Open (Soccer) Science can increase the cognitive demands of doing good science by accentuating the difficulties of finding relevant knowledge in the sea of available information, assessing the quality of contributions, connecting pieces of knowledge across contributions, and tracking the changing status of claims. What cognitive strategies will scientists develop to operate effectively in an Open Science world? What strategies will the scientific community, as a whole, develop? How will Soccer Science be played?

4. A platform in support of Open Science exploration

Many sorts of work practices and tools might moderate the problems introduced by the conflicting cognitive demands of efficient sharing and reliable analysis. Efficient sharing
can be supported if, when distributing the information to the community, scientists use the same external representations that they developed in the course of their own private reasoning. A common, sharable representation can reduce communication overhead in moving ideas from the private to public domain by reducing the need for translation. In addition, the cost of information uptake can be reduced if the provenance, reliability, and context of found information is easily accessible. Also, the burden of distribution is reduced if the scientist can easily and directly control when and with whom the local information is distributed or updated.

To explore the practice and tools that might be useful in an Open Science world, we have developed a research platform, called CACHE, that allows us to investigate different modes of community practice and types of interaction that would enable scientists to play Soccer Science productively. In this section we outline three developmental stages of the CACHE platform: ACH0, which supports individual analysts, CACHE-A, which supports small groups collaborating in real time, and CACHE-BC, which supports diverse structures of interaction across a community that may be widely dispersed in time and space, and which may not even be part of an explicit collaboration. In addition to serving as an experimental platform, CACHE models possible computational support for Open Science on multiple levels: the individual scientist, the interacting group, and the broader community. At the individual level, CACHE incorporates methods to debias individuals in analysis tasks. At a group level, it includes tools for sharing representations of a focal problem and supporting awareness within a working group. At a community level, it allows easy distribution of, and relevance-based access to, new findings.

The cognition of decision makers, including scientists, is affected by biases in making judgments (i.e., systematic errors and limitations). The biases are generally amplified by complex decision tasks that, as for scientific tasks, require chained inferences (Gettys et al., 1973). Corrective interventions or aids such as the introduction of apt representations can reduce some of these biases (e.g., Lim & Benbasat, 1997). Groups of decision makers are also biased to some degree. The quality of group decisions can be improved by properly engineering the properties of the groups (Schultz-Hart, Frey, Lüthgens, & Moscovici, 2000) or the tools that support the decision groups (see Convertino, Billman, Pirolli, Massar, & Shrager, 2008). CACHE incorporates methods for reducing biases and improving decision quality of scientists.

CACHE capitalizes on an analytical method, called ACH—the Analysis of Competing Hypotheses—developed by Richards Heuer (1999) to improve reasoning by intelligence analysts. Heuer noted a series of cognitive biases and shortcuts that can impair intelligence analysis, including biases in the perception of cause and effect, biases in estimating probability, hindsight biases, and others. Our research has focused on “confirmation bias,” or the tendency to be overly committed to an initial hypothesis or belief, despite evidence that should prompt revision: “people do not naturally seek disconfirming evidence, and when such evidence is received it tends to be discounted” (Heuer, 1999, p. 46).

The relations between confirmation bias, science, and tool development merit unpacking. First, adherence to prior belief given conflicting evidence is not always biased: The body of established support may outweigh the new, conflicting information. Indeed, observation of
practicing scientists’ first reaction to a surprising result showed that they first checked whether something was wrong with the method rather than immediately accepting the result at face value (Fugelsang, Stein, Green, & Dunbar, 2004). Bias arises when decision makers tend to err systematically in one direction, not because decision makers are sensitive to prior beliefs. Second, while it is difficult to identify what the “correct” degree of adherence to prior belief may be in the course of doing science, in situations where this can be assessed people systematically tend toward over-, not under commitment to initial beliefs. Third, the tools and methods we advocate should reduce bias of either sort; should there be situations in which scientists are too rash, not just too conservative, our methods should help here as well. Improved reasoning should result from a systematic and extensive consideration of relevant evidence and hypotheses, and this is what our tools are designed to support. Finally, we note that CACHE can also aid in integrating information from multiple attempted replications of a new finding. Confirming findings through replication and tracking the evolving methods around new phenomena are important functions of science (Gorman, 1992), and CACHE can support organized accumulation of evidence of this sort.

Heuer not only notes the occurrence of bias, but prescribes a method for improvement, the Analysis of Competing Hypotheses (ACH). ACH provides a systematic process for evaluating hypotheses in light of evidence. The ACH method requires that the analyst fill out a decision matrix, where the rows represent evidence and the columns represent mutually exclusive and covering hypotheses (Fig. 1, and Heuer, 1999, p. 97). The analyst is encouraged to fill in each cell of the matrix to indicate how each piece of evidence (row) bears on each hypothesis (column), indicating, for example, that the evidence is “consistent,” “inconsistent,” or “neutral” with respect to the hypothesis. The evidence can be integrated to decide among alternatives. The ACH method is based on classical decision matrices, with evidence playing the role of the choice attributes. Pirolli, Good, Heiser, Shrager, and Hutchins (2005) developed a digital implementation of ACH, here called ACH0, and empirically demonstrated that it supports the ACH method at least as efficiently as the “pen and paper” practice of ACH analysis prescribed by Heuer. Pirolli et al.’s electronic ACH0 is beginning to gain adoption in the U.S. intelligence community.

Several characteristics of the basic ACH method deserve note with respect to the cognitive considerations of distributed science discussed above. First, all analysts using the ACH method utilize the same decision-matrix representation, leading to simplicity in sharing. Second, decision matrices are simple to use and understand—they are “mind sized”—and can be used to organize analyses in self-contained, piecemeal, understandable subsets. Third, encouraging analysts to fill out the entire matrix focuses them on disconfirming as well as confirming evidence, and it helps to compensate for systematic biases such as confirmation bias. The ACH method was originally conceived for individual decision makers. However, in CACHE it is also aimed at improving the reasoning of groups and communities. Indeed, the tendency of scientific communities to confirm, rather than falsify theories, or to support the dominant paradigm has been pointed out by many philosophers and historians of science, most famously Kuhn (1962) and Lakatos (1978).

Neither ACH nor its electronic instantiation, ACH0, support collaboration, and useful as the ACH method may be, decision matrices are not an adequate representation for complex
multistage inferences, where the problem is being analyzed in small components distributed across the community. Toward studying these aspects of distributed analysis, Convertino et al. (2008) developed and studied a suite of collaborative ACH tools, collectively called CACHE (Collaborative ACH Environment). CACHE improves upon Pirolli et al.’s ACH0 in a number of ways. First, CACHE is Web-based; multiple analysts can work on different parts of the same problem simultaneously, each incorporating different evidence into their analyses and sharing their analytical matrices with one another. Convertino et al. used an early version of CACHE (CACHE-A) to study how this form of collaboration might influence analytical biases in groups, like those identified by Heuer for individuals. This is of significant concern because, as previously described, collaboration could accentuate or minimize bias; for example, collaborating analysts might inappropriately entrain one another’s analyses, rather than harvesting the benefits of bringing different points of view to a problem.

CACHE-A extends ACH, permitting multiple analysts to collaborate while retaining the “mind-sized-ness” of the ACH matrix, and Heuer’s other insights. CACHE-A also inherits ACH’s clear distinction between evidence and hypotheses. The CACHE platform includes a suite of tools that support real-time collaborative work in teams: A “ticker” notifies when a the analysis of a collaborator has changed, analysts in the same team can then view one another’s matrices, and they can use a chat tool to communicate about the analysis.

Fig. 1. Two decision matrices are depicted “owned” by different users (Heuer and Shrager), engaged in different aspects of the analysis of a crime. The matrices are linked by a hypothesis → evidence provenance edge. Presumably Heuer created his matrix and then promoted the linked hypothesis to the assertion pool. Shrager discovered this promoted assertion and used it as evidence in his analysis. The resulting provenance connection is explicit from the point of view of the community, and yet implicit to the individual users.
5. From real-time group collaboration to the Bayes community: CACHE-A to CACHE-BC

CACHE-A extended ACH0 to enable real-time collaboration in teams, but, as we have noted, science is a collaboration that extends not just across individuals and teams but also across loosely coupled communities distributed in space and time. In “Football Science” time is less of the essence as there is time to consider one’s options, but Open (“Soccer”) Science will require effective support for collaborations taking place much more rapidly—at what we shall call “Web Speed.” Web Speed collaboration may well be worse than either real-time (e.g., face-to-face, or “on the phone”) or long-term (e.g., via email) collaborations. In real-time collaborations, the participants are working as a unit—able to immediately work out problems such as misunderstandings. Similarly, in the long, slow collaborations that characterize Football Science, there is sufficient time to reflect and try out a number of combinations if problems arise. At Web Speed one neither has the advantages of being able to interact with the immediacy of real time, nor of having the time to reflect. The CACHE Bayes community (CACHE-BC) model was designed with this problematic balance in mind.\(^2\)

CACHE-BC handles real-time revision updating through explicit representation of evidential provenance. The conclusions in a given study, here, as before, represented by ACH matrices, are generally built on preexisting conclusions from other studies (matrices). If the latter are revised, the former should be updated. Recall in our vignette about Anne, Barry, and Chris that when Chris’s new analysis led to a revision in Barry’s analysis of the function of Gb, Anne was unable to rapidly (if at all) find out about this change. In these situations, a computational infrastructure for real-time revision updating would enable the automatic propagation of information up the evidential chain. CACHE-BC enables the community to work on many separate, but interrelated problems and to have real-time revision updating so that everyone who utilizes a piece of information can find out when any factors relating to that information, such as its content, quality, or basis, change. The way that this works in CACHE-BC is through the explicit representation of evidential provenance: When Anne bases an analysis on Barry’s analysis, the linkage between these is explicitly maintained, so that when Chris leads Barry to update his result, Anne is automatically notified. Further, everyone who had depended upon Anne’s result would be recursively notified of this change, so that they could reconsider their analyses. This simple mechanism leads to a very powerful new way of playing Soccer Science.

CACHE-BC represents a significant conceptual advance on both Heuer’s ACH method (and its implementation: ACH0) and on CACHE-A. In both of those cases, the ACH matrix representation both affords and enforces a clear distinction between evidence and the hypotheses: evidence in rows, hypotheses in columns. In constructing an ACH matrix the analyst selects evidence from a central pool (here called the “assertion pool”) and assigns these to rows in the ACH matrix. Analysts cannot add to or delete assertions from the pool itself nor modify individual assertions in the pool. Put in another way, the public state of belief in evidence in the ACH world (i.e., the community’s information space) is not itself
subject to analysis; specific analysts can include or exclude evidence from their own analyses and differentially weight these in a given analysis, but analysts cannot change the evidence; the assertion pool is static; with respect to the community of reference, evidence exists, so to speak, “in the wild.” This is a critical assumption of both ACH and CACHE-A: The value of evidence differs only in the context of particular analyses, and a particular analysis is precisely what is represented by a particular ACH matrix.

Real-time revision updating through explicit representation of evidential provenance is supported by a key property of CACHE-BC, which we call “promotion.” Whereas ACH0 did not permit any form of collaboration, and CACHE-A permitted analysts to examine one another’s analyses within a group, CACHE-BC supports the broader function for which one analyst’s hypothesis can be used by another analyst, who came afterwards, as part of his evidence. This is accomplished simply by allowing analysts to “promote” their hypotheses to the public assertion pool. Promoted hypotheses, from that point onward, appear to all analysts just like evidence: They can be selected as evidence by any other analyst in the community and added as evidence to any analytic matrix (Fig. 2).³⁴ Promotion is, in CACHE-BC, the analog of publishing a new result in Football Science, or, for example, blogging a result in Soccer Science; once promoted, hypotheses are available for use as evidence in other analyses conducted by any other analyst.

Allowing hypotheses to be promoted to the assertion pool and then used as evidence in other analyses qualitatively changes the collaboration process. To see this, let us

![Diagram](image-url)
return to Anne, Barry, and Chris: Suppose that Anne, Barry, and Chris were all using CACHE-BC instead of merely publishing in standard blogs. Publication (promotion) is still immediate, just as in the typical blog world: When Barry promotes his hypothesis that Gb functions as glucokinase, it is placed into the assertion pool. Anne accesses Barry’s promoted hypothesis by importing it as evidence into the matrix she is using to analyze Ga. She concludes (hopefully through other evidence as well) that Ga functions as a glucokinase as well. She likewise promotes her hypothesis, which others import as evidence into their analytical matrices, and so on. When, later, Chris draws Barry’s conclusion into question, Chris will (one hopes) promote this result, inform Barry, who will presumably import Chris’s new conclusion into his (Barry’s) Gb analytical matrix. Normally, that is without CACHE-BC, Barry’s change at this point would not affect Anne’s conclusions, which, after all, were made in the past. Furthermore, again without CACHE-BC, anyone who relied upon Anne’s conclusions would also not be able to find out about the change in Barry’s analysis of Gb (unless they, and Anne, were highly attentive to the literature). But because CACHE-BC exhaustively retains the explicit linkages between Barry’s hypothesis about Gb and Anne’s use of it, and between Anne’s hypothesis about Ga and others’ uses of this, Anne, and others who use her results can be informed of potentially important changes in results that they have relied upon. Once linked up in this way, CACHE-BC can track changes and possibly update confidences through to all of the conclusions that may be affected by those changes.5,6

As a result of this facility, complex problems can, in CACHE-BC, be broken down into overlapping subproblems, and these can be distributed throughout the community of decision makers. Different matrices may overlap in terms of the evidence they rely upon, or in terms of the hypothesis under consideration, and may represent subgoal/supergoal or prior/posterior relationships with respect to one another, or, as in Convertino et al., different ways of relating the same evidence to the same or similar hypotheses. Indeed, one would expect the community to almost always divide up the problem space into subcomponents through their own self-selection of problems to work on and evidence to use in analysis; CACHE-BC is merely a device for provenance tracking: recording the subjective relationships between evidence and hypotheses, and the logical dependencies among contributions across a community of analysts distributed in space and time. Once one has such a record, it can be utilized to increase the efficiency of analysis by the community in the ways described and probably in other ways as well.

It is widely understood that tracking provenance is important, but CACHE-BC goes beyond a simple scientific citation tree—the directed acyclic graph of papers citing other papers—by stating how particular hypotheses rely upon particular evidence. In addition to permitting the sort of updating described above, the CACHE-BC model of provenance tracking helps to increase the signal-to-noise ratio in the face of very large and noisy knowledge bases, partially ameliorating the problems in our conflicting gene function example, as well as in cases of noisy data. This sort of technology simplifies revision updating, marking results as questionable, informing scientists when results that they have depended upon come into question, and “chasing” such revisions through the provenance tree.
6. Discussion

The line of work reviewed in this paper has led us from a pen-and-paper technique for overcoming cognitive bias in analysis (ACH: Heuer, 1999) and a direct digital implementation of that technique (ACH0: Pirolli et al., 2005), to an extension of that technique supporting simultaneously collaborating teams of analysts (CACHE-A: Convertino et al., 2008), and finally to a way of enabling the management of evidence and hypotheses across multiple stages of inference that cross both communities and time (CACHE-BC).

Our goal in this research program has been twofold: first, to articulate a design rationale and the development of systems that can support multistage reasoning at the individual, group, and community levels, and second to provide a means of empirically investigating new forms of societal organization in analytical (esp. scientific) collaboration. The process of developing and testing the various instantiations of CACHE drew on, and in turn increased our understanding of collaborative analysis tasks, which have not yet been investigated extensively by the distributed cognition community. We believe, following Kraemer and King (1988), that for decision support designs to succeed, the development of technology and the evaluation of its impact on decision processes need to be combined.

CACHE-A and -BC both allow individual analysts to maintain their private, local, perspectives on relatively simple subproblems while, at the same time, allowing them to collaborate on complex problems. CACHE-BC goes beyond CACHE-A by enabling analysts to depend upon one another’s results and to depend upon the efficient updating of belief as new evidence is collected. This model may free analysts from worrying about the risk that the knowledge used now could become outdated or inaccurate in the future, because they know that its quality will be updated dynamically as provenance is automatically maintained by the system; one is not required to think at all about the structure of the provenance network, thus reducing cognitive load and freeing resources. CACHE thereby enables the community to integrate large amounts of heterogeneous information, share hypotheses and interpretations, discover relevant evidence, and maintain provenance linkages between stages in the multistage process.

We developed the CACHE series of experimental platforms in order to explore more efficient forms of knowledge sharing and multistage reasoning in scientific communities. The platforms, and especially CACHE-BC, honor to various degrees the cognitive principles described at the beginning of this paper. Although initially motivated by considering the work of individual scientists, these principles are also (if not more) applicable to work of interacting groups, or indeed to the whole community of scientists where the individual’s ability to be successful is dependent on coordinated cognitive interactions with collaborators and the cognitive products that these interactions generate. Thus, CACHE-BC should simultaneously provide support at the levels of the individual scientist, the group, and the community. At the individual level, it offers tools to support foraging for new information and debiasing individuals in analysis tasks. At a group level, it includes tools for sharing representations of a focal problem and for improving awareness of judgments about the shared evidence pool. At a community level, it supports the functions of provenance tracking and automatic updating, as various findings relevant to individuals or groups are revised.
across the community. Finally, CACHE-BC’s support for problem decomposition into sub-
problems (i.e., as interconnected decision matrices) and for collaborative multistage reason-
ing across these subproblems reduces cognitive costs at each of the levels of analysis.

To summarize, let us return to the five principles described in the introduction and briefly
recount how CACHE addresses them.

6.1. ‘‘Mind-sized’’ analytical sets

All of the ACH-line of methods hark to mind-sizedness, but CACHE-BC especially
addresses this principle because complex problems can be broken down into mind-sized
matrices, constructed by one or a small number of collaborators, that refer to one another. In
this way a complex problem is divided (and hopefully conquered) through many inter-
referred mind-sized chunks.

6.2. Apt representation

The ACH-line of methods provides a common meta-representation in the matrices for the
relation between evidence and hypotheses. This is not, of course, a complete resolution to
the problem of apt representation among hypotheses, but it is a start. The latter is a much
more difficult problem, pervading all technologies of this sort, and is not directly addressed
by any of the ACH methods.7

6.3. Efficient and controlled sharing

Both CACHE-A and CACHE-BC carefully manage the sharing of information, albeit in
very different ways. In a sense, this is the central point of the whole line of CACHE meth-
ods, and it is where most of the important future work will take place. (See the section on
future work, below.)

6.4. Task orientation

The ACH-line of methods organizes information for any of several tasks into its own
interrelated set of matrices. This organization facilitates working on and switching among
multiple tasks, as typically needed by scientists. Switching to a new task by switching to a
new matrix set provides the scientist with an integrated representation of the task, preserved
its context, and thus allows a scientist easily to return to an ongoing task.

6.5. Multistage inference

This is, of course, CACHE-BC’s raison d’etre. Not only do these methods support indi-
viduals reasoning over complex evidence chains, they also support reasoning distributed
across people and time. Public and up-to-date records for each inference allow the members
to revise and build upon prior inferences.
7. Relation to prior work

Our interest in supporting the dynamics of Open Science is shared by other researchers. In particular, prior work has emphasized the importance of efficient sharing. An early example is ArXiv.org, a dramatically successful project for sharing well-developed results (physics preprints) in an electronic archive. Developed in 1991, physicists are now expected to upload their papers simultaneously with sending them to journals for official publication, and thousands of e-prints are added every month. Thus, the worldwide community has immediate free access to novel results. More recent Web-based capabilities such as social networking, social annotations, wiki applications, and semantic Web techniques portent major transformations in how scientists share and collaborate (e.g., Bradley et al., 2008).

Citeseer was the first Web-based system to introduce automated citation indexing and citation linking and includes automatic notifications of new citations and recommendations of new papers based on a user profile. Bollacker, Lawrence, and Giles (1998) initially developed the system as a research prototype, but it has become a highly successful public search engine. Recently, Farooq, et al. (2008) explored novel collaborative features that could support a collaborative laboratory around Citeseer: social networking tools, activity awareness tools, and notifications. This work was inspired by prior research on scientific collaboratories (Finholt & Olson, 1997). Citeseer focuses on provenance maintenance, but not reasoning over the information accessed. Indeed, we do not know of other tools that support the inference integration aspect of distributed multistep inference, though the need was originally noted by Johnson and Halpin (1974).

Looking farther afield from Open Science, collaborative systems developed for the workplace have suggested the cost-saving benefits of a shared workspace that organizes the sharing and the collaboration around the task (or shared activity). Examples include the IBM Activity Explorer prototype and the activity-centric design (Moran, Cozzi, & Farrell, 2005) and the task management augmenting e-mail clients (Bellotti, Thornton, Chin, Schiano, & Good, 2007). Unlike CACHE-BC, these tools do not pertain to collaborative sense making and are not directed at communities.

In short, a variety of successful projects suggest the importance and tractability of supporting Open Science. However, tools that support the multiple factors affecting cognition in Open Science are rare. Our work gives an initial characterization of five important factors, and it provides a tool designed to give multifaceted support.

8. Future directions

Our long-term vision for the CACHE series of platforms is to create a living laboratory enabling us to simultaneously support and study Open Science. We describe some examples of projects toward this end that use the CACHE research platform to explore alternative collaborative practices and policies for scientific communications.

The CACHE matrix provides a cognitively motivated, mind-sized format, one that may contribute to balanced integration of evidence and hence bias reduction. As part of further
empirical explorations of debiasing, it will be important to investigate different evidence integration methods. For example: Under what conditions are assessments improved by weighing evidence by importance or credibility? Is it useful to include irrelevant but discounted evidence, as a means of tracking its status? How does sensitivity to reputation or judged reliability influence assessment, and how good must those assessments be to improve the result?

A second set of questions concern collaboration policies, such as the role of reputation or the speed of notification. Without some careful consideration of such policies, one could be overwhelmed by the number of minor updates that were propagated through the network (as will be very familiar to anyone who uses Twitter, http://www.twitter.com). In our experiments so far we have emphasized the value of notification of changed evidential status, but additional information such as reputation can also be propagated. This might play the role of peer review, potentially acting at the level of both the individual (e.g., “people say this person does good work”) and the evidence (e.g., “people say this study is well done”). More broadly, the effects of equal versus differentiated impact from different sources can be investigated, either in experimentally established or naturally occurring communities. Importantly, the effects of different policies with different assumptions about information quality can be explored through simulation.

Similar issues arise about policies regarding the speed of propagation of updates. To date we have emphasized the potential value of very rapid transmission of information to a wide audience. However, there can also be advantages to slower or more filtered distribution. This might be accomplished by any number of methods, for example, initial propagation to a smaller group, whose response influences the later rate and scope of further distribution. Alternatively, this might be done by system-wide “lag” in distribution, for example, requiring some delay between an author’s initial posting (into a locally viewable workspace) and distribution to wider audiences. Again, different policies of screening could be added to the delay. Finally, to reduce distraction and information load, the notifications could be provided via peripheral displays and digest deliveries that are controlled by the end-user selection rules. Research will be needed on sound policy and practice for using the tools as well as for tool development.

Both empirical studies and simulation in all these areas will be informative, and they could lead to improved support for Open Science based upon the CACHE model.

Notes

1. Provenance refers to the chain of information relating to the origins of something. The term is commonly used in the world of art and artifacts, referring to the documentation that, for example, validates a piece as an original da Vinci (see: http://en.wikipedia.org/wiki/Provenance, accessed: 24 May 2009). In science, provenance could include information about the lab and researchers advancing a claim, the date of discovery, the instruments, methods, models, or analyses used, and so on.
2. We use the term “Bayes” here by reference to the well-known Bayesian method of updating probabilities wherein a posterior probability is calculated via a simple
equation, from prior probabilities and evidence. This method affords iteration over time, as is commonly observed in science, and has been applied in many useful ways, including to networks of hypotheses, which are similar to the model discussed here. Indeed, one way to think of a Bayes community, is as a Bayesian network dispersed across a community of analysts.

3. They can also be selected as hypotheses, enabling multiple matrices to contribute to the support or refutation of given hypotheses.

4. In interface terms, promotion is an almost trivial change in functionality to the tool: A “promote” button appears with each hypothesis.

5. Although the story so far has been linear for simplicity’s sake—one hypothesis arising from, and one piece of evidence entering into each matrix—ACH matrixes will generally import multiple pieces of evidence, and will generally export multiple hypotheses (as Figs. 1 and 2); the appropriate graphical model, therefore, is a directed (acyclic) graph, not the simpler linear graph. This, however, presents no theoretical problems for the proposed upward propagation of the updates.

6. In practice, CACHE can integrate uncertainty in a number of ways, including Bayesian, Dempster-Schafer, continuous or discrete averaging, winner-take-all (actually logit), or spreading activation. The default propagation algorithm is just to notify everyone up the tree that something in the incoming provenance root system has changed. The spreading activation model is like the default propagation algorithm except that it reduces the strength of the notification according to the distance from the change. This has both a discrete and a continuous mode.

7. The CACHE-BC implementation actually addresses this more directly in ways that are too complex to go into here, and which we intend to address in detail elsewhere. Briefly, all assertions in the pool are connected to a semantic network so that one can conduct semantic search for related hypotheses, regardless of the particular vocabularies used by the authors.

Acknowledgments

This work was funded, in part, by Office of Naval research Contract N00014-02-C-0203 to P. Pirolli and S. K. Card, and by Advanced Research and Development Activity Novel Intelligence from Massive Data contract MDA904-03-C-0404 to P. Pirolli and S. K. Card. Stu Card, Ed Chi, Richards Heuer, and Mark Stefik were instrumental in helping us think through some of the subtle issues underlying this work. Mike Gorman and Caroline Verdier shepherded the paper through many reviews. Most importantly, three anonymous reviewers gave generously their time in reviewing the paper several times in great detail and provided many insightful comments. Even in the face of the unstoppable tidal wave of Open Science, we authors are unanimous in our appreciation for reviewers who put so much work into a paper that is not even theirs. They, and many other reviews over the years, have saved us from what would have been countless cases of public humiliation. Additionally, we would like to thank Rebecca Agin for figure art.
References


