Collaborative Research:
CT-T:
Explainable Policies for Establishing Trust in Web Applications

Team Proposal to National Science Foundation in response to Cyber Trust (CT) solicitation NSF-05-518

NSF Proposal number 0524842 submitted February 7 2005

Florida Institute of Human and Machine Cognition
PI: Jeff Bradshaw
co-PI: Pat Hayes

Brigham Young University
PI: Kent Seamons

Stanford University Knowledge Systems Laboratory
PI: Richard Fikes
co-PI: Deborah McGuinness

University of Illinois at Champaign-Urbana
PI: Marianne Winslett

COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE/if not in response to a program announcement/solicitation enter NSF 04-23

NSF 05-518 **02/07/05**

FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.)

CNS - CYBER TRUST

FOR NSF USE ONLY

NSF PROPOSAL NUMBER

0524716

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
				158995659	

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)

200760849

SHOW PREVIOUS AWARD NO. IF THIS IS
☐ A RENEWAL
☐ AN ACCOMPLISHMENT-BASED RENEWAL

IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES ☐ NO ☒ IF YES, LIST ACRONYM(S)

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE

Florida Institute for Human and Machine Cognition, Inc.

AWARDEE ORGANIZATION CODE (IF KNOWN)

6250009905

ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE

40 S. Alcaniz St.
Pensacola, FL 32502-6008

NAME OF PERFORMING ORGANIZATION, IF DIFFERENT FROM ABOVE

PERFORMING ORGANIZATION CODE (IF KNOWN)

ADDRESS OF PERFORMING ORGANIZATION, IF DIFFERENT, INCLUDING 9 DIGIT ZIP CODE

IS AWARDEE ORGANIZATION (Check All That Apply) (See GPG II.C For Definitions)
☐ SMALL BUSINESS ☐ MINORITY BUSINESS
☐ FOR-PROFIT ORGANIZATION ☐ WOMAN-OWNED BUSINESS

☐ IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE

TITLE OF PROPOSED PROJECT **Collaborative Research: CT-T: Explainable Policies for Establishing Trust in Web Applications**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ **887,652**	**36** months	**09/01/05**	

CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW

☐ BEGINNING INVESTIGATOR (GPG I.A)
☐ DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C)
☐ PROPRIETARY & PRIVILEGED INFORMATION (GPG I.B, II.C.1.d)
☐ HISTORIC PLACES (GPG II.C.2.j)
☐ SMALL GRANT FOR EXPLOR. RESEARCH (SGER) (GPG II.D.1)
☐ VERTEBRATE ANIMALS (GPG II.D.5) IACUC App. Date _____
☐ HUMAN SUBJECTS (GPG II.D.6)
Exemption Subsection _____ or IRB App. Date _____
☐ INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED (GPG II.C.2.g.(iv).(c))
_____
☐ HIGH RESOLUTION GRAPHICS/OTHER GRAPHICS WHERE EXACT COLOR REPRESENTATION IS REQUIRED FOR PROPER INTERPRETATION (GPG I.E.1)

PI/PD DEPARTMENT

PI/PD POSTAL ADDRESS
40 South Alcaniz Street

Pensacola, FL 32502
United States

PI/PD FAX NUMBER
850-202-4440

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Electronic Mail Address
PI/PD NAME **Jeffrey M Bradshaw**	**PhD**	**1996**	**850-232-4345**	**jbradshaw@ihmc.us**
CO-PI/PD **Patrick J Hayes**	**PhD**	**1973**	**850-202-4416**	**phayes@ihmc.us**
CO-PI/PD				
CO-PI/PD				
CO-PI/PD				

Electronic Signature

CERTIFICATION PAGE

Certification for Authorized Organizational Representative or Individual Applicant:

By signing and submitting this proposal, the individual applicant or the authorized official of the applicant institution is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding debarment and suspension, drug-free workplace, and lobbying activities (see below), as set forth in Grant Proposal Guide (GPG), NSF 04-23. Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, Section 1001).

In addition, if the applicant institution employs more than fifty persons, the authorized official of the applicant institution is certifying that the institution has implemented a written and enforced conflict of interest policy that is consistent with the provisions of Grant Policy Manual Section 510; that to the best of his/her knowledge, all financial disclosures required by that conflict of interest policy have been made; and that all identified conflicts of interest will have been satisfactorily managed, reduced or eliminated prior to the institution's expenditure of any funds under the award, in accordance with the institution's conflict of interest policy. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Appendix C of the Grant Proposal Guide.

Debarment and Suspension Certification

(If answer "yes", please provide explanation.)

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency? Yes ☐ No ☒

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Appendix D of the Grant Proposal Guide.

Certification Regarding Lobbying

This certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding $100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding $150,000.

Certification for Contracts, Grants, Loans and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

(1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, ''Disclosure of Lobbying Activities,'' in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, Title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than $10,000 and not more than $100,000 for each such failure.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME **Larry L Warrenfeltz**	**Electronic Signature**	**Feb 7 2005 5:44PM**

TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS	FAX NUMBER
850-202-4473	**lwarrenfeltz@ihmc.us**	**850-202-4540**

*SUBMISSION OF SOCIAL SECURITY NUMBERS IS VOLUNTARY AND WILL NOT AFFECT THE ORGANIZATION'S ELIGIBILITY FOR AN AWARD. HOWEVER, THEY ARE AN INTEGRAL PART OF THE INFORMATION SYSTEM AND ASSIST IN PROCESSING THE PROPOSAL. SSN SOLICITED UNDER NSF ACT OF 1950, AS AMENDED.

Project Summary.

As daily reliance on networked information and services becomes increasingly vital to every aspect of our lives, the requirement for dependable functionality of both national infrastructures and localized embedded systems naturally moves to the forefront. These systems must be dependable and trustworthy even in the face of cyber attacks and periodic fluctuations of capability. People need assurance of the integrity and confidentiality of organizational and personal information, and to have confidence that system behavior will conform to policy constraints; but the network must be capable of responding adequately to important and unusual information requests.

This requires research at several levels. Information exchange protocols must be secure and scalable, able to deal with issues of information release and exposure, credential checking to establish authenticity of remote participants on an open network, and attribution and reliability and timeliness of information. Software agents must behave in ways that are reliably sensitive to policies concerning obligations, constraints and permissions; software must be capable of reasoning about such policies and applying them reliably, and finally, a working system must be capable of communicating explanations of its decisions to human users in a comprehensible manner, and able to support after-the-fact analyses of why decisions were taken and what reasoning justified them. Researchers from psychology, sociology, artificial intelligence, and computer security have investigated all of these issues within their own domains, but have never tried to integrate their approaches to create an end to end, scalable, human-friendly, and secure architecture for cross-organizational information sharing. This proposal aims to create such a framework and evaluate it in the context of cross-organizational information sharing for disaster response in Champaign, Illinois.

The research threads being woven together to create the framework are the TrustBuilder project (BYU and Illinois) for scalable, policy-based trust establishment in virtual organizations; the KAoS project (IHMC) for policy-based agent reasoning; the Inference Web project (Stanford KSL) for web-based handling of justifications proofs and generation of explanations; and the Common Logic project (IHMC) which supplies a semantically secure basis for a wide variety of kinds of information and reasoning. In addition, IHMC researchers will study the influence of cognitive bias on human understanding of explanations under conditions of stress. These all represent different levels of analysis, from controlling details of safe Web transactions up to human perception of trustworthiness of an entire system, or even an entire technology. This proposal forces an integration of these different levels of analysis together on a concrete and demanding application, with the ultimate goal of developing of a unifying *theory* of trust negotiation and inference which can be applied as a single connecting framework across all the technological levels present in virtual organizations.

Intellectual Merit. By pursuing an interdisciplinary research thread that begins in basic theory, is applied in the context of challenging elements of trustworthy system components, is deployed in realistic use scenarios, and is extended to achieve a better understanding of social mechanisms relating to trust, the proposed work will expand our understanding of trustworthy systems and their place in society in a way that single-discipline efforts cannot.

Broader Impact. The work will involve student effort at the undergraduate, master's, doctoral and postgraduate levels at four institutions, with mutual coordination between them. Insights and ideas which emerge will be incorporated rapidly into teaching curricula at BYU and Illinois, and (through other linked collaborative activity) in Southampton. Two of the co-PIs are women who exhibit leadership in IT research and are role models for female students. IHMC will support high school student research assistants, and IHMC faculty and scientists will participate in monthly science-oriented programs for elementary school age children. In the longer term, the project's results have the potential to improve disaster response in small- and medium-size cities across the nation. The PIs will continue their long-standing participation in world-wide standard-setting activities, including W3C Working Groups and ISO standardization efforts, so that technical results and insights will be quickly incorporated into emerging standards. Finally, software produced by the project will be made available for public use.

C. Project Description

C 1. Introduction and overview

As daily reliance on networked information and services becomes increasingly vital to every aspect of our lives, the requirement for dependable functionality of both national infrastructures and localized embedded systems naturally moves to the forefront. These systems must be dependable and trustworthy even in the face of cyber attacks and periodic fluctuations of capability. People need assurance of the integrity and confidentiality of organizational and personal information, and to have confidence that system behavior will conform to policy constraints; but the network must be capable of responding adequately to important and unusual information requests.

This requires research at several levels. Information exchange protocols must be secure, able to deal with issues of information release and exposure, credential checking to establish authenticity of remote participants on an open network, and attribution of reliability and timeliness to information. Software agents must behave in ways that are reliably sensitive to policies concerning obligations, constraints and permissions; software must be capable of reasoning about such policies and applying them reliably, and finally, the entire system must be capable of communicating explanations of its decisions to human users in a comprehensible manner, and able to support after-the-fact analyses of why decisions were taken and what reasoning justified them. Various research projects are under way in all these areas, but in relative isolation from one another. Researchers from psychology, sociology, artificial intelligence, and computer security have investigated all of these issues within their own domains, but have never tried to integrate their approaches to create an end to end, scalable, human-friendly, and secure architecture for cross-organizational information sharing. This proposal aims to create such a framework and evaluate it in the context of cross-organizational information sharing for disaster response in Champaign, Illinois.

The main research threads being woven together here are: automated trust negotiation, exemplified in the TrustBuilder project (BYU and Illinois) which defines exchange protocols for secure information transfer and resource access on the Web; the KAoS project (IHMC) which uses Web ontology standards to perform complex reasoning about policies, delegations and agent actions; the Inference Web project (Stanford KSL) which provides a notation and framework for transmitting proofs on the Web and generating explanations from them; and the Common Logic project (IHMC) defining a highly expressive logical framework proposed for ISO standardization. Several subsets of these projects have a proven track record of successful cross-country collaboration with one another. Other team members supply related expertise from other perspectives: in particular, Paul Feltovich (IHMC) will consider the ways that cognitive bias influences the human perception of explanations under conditions of stress. Several of the PIs and co-PIs are actively involved in semantic web standardization efforts, providing another unifying theme. Part of our methodology is to utilize current and planned standard languages and methods, such as W3C recommendations and ISO standards, as far as possible, and we expect this work to result in recommendations for improved and future standards

All the participants are interested in, and committed to, the evolution of a unifying *theory* of trust negotiation and inference which can be applied as a single connecting framework across the various technological levels. Though in some cases research on point solutions addressing focused subsets of the issues is underway, no such framework yet exists which can be adequately applied across all the layers of a full Web trust architecture. Other overarching goals are to isolate and characterize the kinds of reasoning that are required for realistic trust and policy use in heterogeneous and distributed networks, with a view to designing optimized trust reasoners, and to critically examine the ways that human confidence in a network may be influenced by cognitive biases.

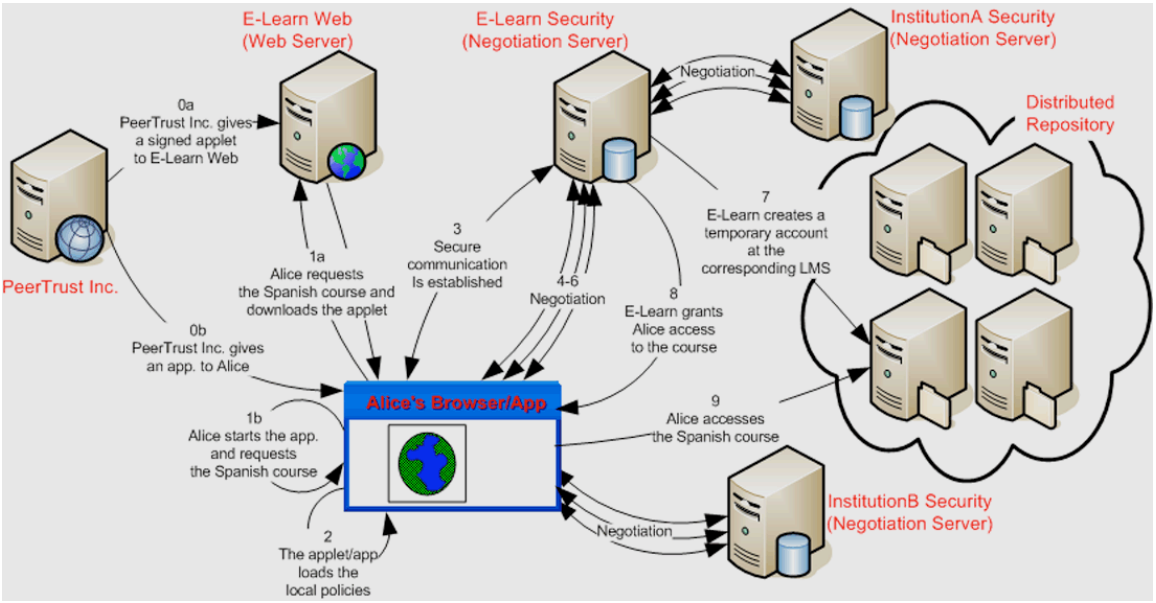
C 2. States of the Arts.

C 2.1 Automated Trust Negotiation (Seamons and Winslett, linked projects from BYU and Illinois)

TrustBuilder, a joint project of BYU and Illinois, is a framework for trust negotiating in network transactions, focusing particularly on techniques for secure implementation of policies controlling Web information access. Trust negotiation differs from traditional identity-based access control and information release systems mainly in the following aspects:

1. Trust between two strangers is established based on parties' properties, which are proven through disclosure of digital credentials.
2. Every party can define access control and release policies (policies, for short) to control outsiders' access to their sensitive resources. These resources can include services accessible over the Internet, documents and other data, roles in role-based access control systems, credentials, policies, and capabilities in capability-based systems. The policies describe what properties a party must demonstrate (e.g., ownership of a driver's license issued by the State of Illinois) in order to gain access to a resource.
3. Two parties establish trust directly without involving trusted third parties, other than credential issuers. Since both parties have policies, trust negotiation is appropriate for deployment in a peer-to-peer architecture such as the Semantic Web, where a client and server are treated equally. Instead of a one-shot authorization and authentication, trust is established incrementally through a sequence of bilateral credential disclosures.

A trust negotiation process is triggered when one party requests to access a resource owned by another party. The goal of a trust negotiation is to find a sequence of credentials ($C1, \ldots, Ck, R$), where R is the resource to which access was originally requested, such that when credential Ci is disclosed, its policy has been satisfied by credentials disclosed earlier in the sequence or to determine that no such credential disclosure sequence exists. The overall structure of the process can be gleaned from this diagram, reproduced from the PeerTrust website, http://www.learninglab.de/peertrust/.

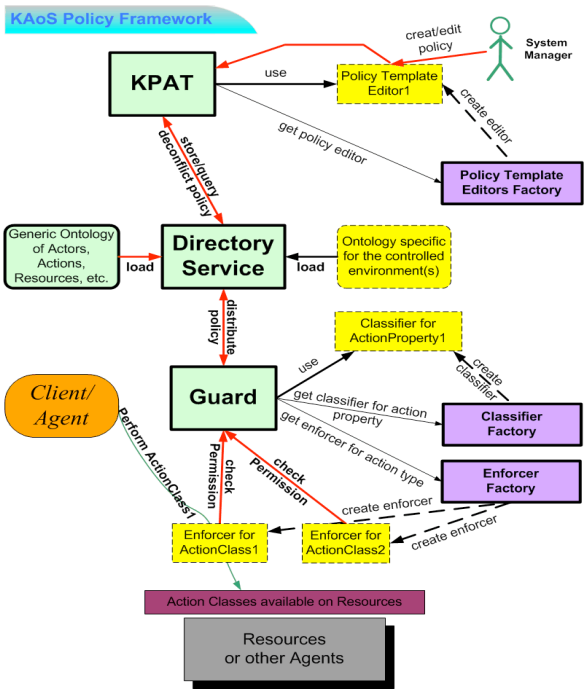


TrustBuilder is a relatively mature software component; for example, a concurrent effort is adapting the TrustBuilder prototype software for trust establishment for use within the Grid Security Infrastructure. So far, however, it is limited in its use of formal ontologies and techniques for auditing and explanation, a limitation that we propose to address.

C 2.2 KAoS (Bradshaw, IHMC)

KAoS is a collection of componentized services compatible with several popular agent platforms, including the DARPA CoABS Grid [UBJ04a], the DARPA ALP/Ultra*Log Cougaar agent framework (*http://www.cougaar.net*), CORBA (*http://www.omg.org*) and Brahms [s02]. The adaptability of KAoS is due in large part to its pluggable infrastructure based on Sun's Java Agent Services (JAS) (*http://www.java-agent.org*). While initially oriented to the dynamic and complex requirements of software agent applications, KAoS services have also been adapted to general-purpose grid computing [JCJ+03] and Web Services [UBJ04a] environments.

Under DARPA and NASA sponsorship, we have been developing the KAoS policy and domain services to increase the assurance and trust with which agents can be deployed in a wide variety of operational settings. *KAoS Domain Services* provide the capability for groups of software components, people, resources, and other entities to be semantically described and structured into organizations of domains and subdomains to facilitate collaboration and external policy administration. *KAoS Policy Services* allow for the specification, management, conflict resolution, and enforcement of policies within domains. The figure presents basic elements of the KAoS framework, emphasizing infrastructure supporting the specification and use of authorization policies. There are additional components to support obligation policies and other aspects of the system.



Framework functionality can be divided into two categories: generic and application/platform-specific. The generic functionality includes reusable capabilities for:

- Creating and managing the set of core ontologies;
- Storing, deconflicting and querying;
- Distributing and enforcing policies;
- Disclosing policies.

For specific applications and platforms, the KAoS framework can be extended and specialized by:

- Defining new ontologies describing application-specific and platform-specific entities and relevant action types;
- Creating extension plug-ins specific for a given application environment such as:

- Policy Template and Custom Action Property editors;
- Enforcers controlling, monitoring, or facilitating subclasses of actions;
- Classifiers to determine if a given instance of an entity is in the scope of a given class-defining range.

KAoS uses ontology concepts (encoded in OWL) to build policies. During its bootstrap, KAoS first loads the core KAoS Policy Ontology defining concepts used to describe a generic actors' environment and policies in this context (http://ontology.ihmc.us). Then, KAoS loads additional ontologies on top of this, extending concepts from the core ontology, with notions specific to the particular controlled environment and application domain.

The KAoS Policy Service distinguishes between *authorizations* (constraints that permit or forbid some action) and *obligations* (constraints that require some action when a state- or event-based trigger occurs or that serve to waive such a requirement). Other policy constructs (for example, delegation or role-based authorization) are built from the basic domain primitives plus four policy types.

KAoS policy's OWL definition is an instance of one of these four basic policy classes: PositiveAuthorization, NegativeAuthorization, PositiveObligation, or NegativeObligation. The property values determine management information for a particular policy (for example, its priority). The type of policy instance determines the kind of constraint KAoS should apply to the action, while a policy's action class is used to determine a policy's applicability in a given situation. An action class helps classify action instances that actors intend to take or are currently undertaking. Components (such as KAoS Guards) that are interested in checking policy impact on these actions construct RDF descriptions of action instances. KAoS classifies these instances, relying on the inference capabilities of Stanford University's Java Theorem Prover (JTP, www.ksl.stanford.edu/software/JTP), which will facilitate integration with Inference Web. It then obtains a list of any policies whose action classes are relevant to the current situation. In the next step, KAoS determines the relative precedence of the obtained policies and sorts them accordingly in order to find the dominating authorization policy. If the dominating authorization is positive, KAoS then collects, in order of precedence, obligations from any triggered obligation policies. KAoS returns the result to the interested parties—in most cases, these parties are the enforcement mechanisms that are jointly responsible for blocking forbidden actions and assuring the performance of obligations.

Representing policies in OWL facilitates reasoning about the controlled environment, policy relations and disclosure, policy conflict detection, and harmonization. It also facilitates reasoning about domain structure and concepts exploiting the description logic subsumption and instance classification algorithms. KAoS can identify and, if desired, harmonize conflicting policies through algorithms that we have implemented in JTP. KAoS is a mature project. over the past few years, KAoS services have been used in conjunction with a wide range of applications and operating platforms.

C 2.3 Inference Web (McGuinness, Linked project from KSL-Stanford)

The Inference Web (IW) [McP04a, McP03] aims to take opaque query answers and make the answers more transparent by providing explanations. Inference Web provides an infrastructure for providing explanations from distributed hybrid question answering systems. It utilizes a proof Interlingua – the Proof Markup Language (PML) [PMF05] to encode justifications of information manipulations. It also provides numerous services for manipulating PML documents. It includes a browser for viewing information manipulation traces, an abstractor for rewriting PML documents so that the low level machine-oriented proofs can be transformed into higher level human-oriented explanations, and an explainer to interact with users by presenting explanations and corresponding follow-up questions. It also includes services for helping question answering systems to generate PML, check PML documents for valid applications of inferences, and services for automatic registration of sources and meta-information. The explanations include information concerning where answers came from and how they were derived (or retrieved). The Inference Web infrastructure also includes an extensible web-based registry [PMM03] containing details on information sources, reasoners, languages, and rewrite rules. Source information in

the IW registry is used to convey data provenance. Representation and reasoning language axioms and rewrite rules in the IW registry are used to support proofs, interoperability, and proof combination. The IW browser is used to support navigation and presentations of proofs and their explanations. The explainer is used as an interface to provide a multitude of strategies (such as summaries, graphical depictions, interactive dialogues, etc) for presenting the information.

The Inference Web is in use by several Semantic Web agents using embedded reasoning engines fully registered in the IW. These engines include first order logic reasoners such as Stanford's JTP engine and SRI's SNARK engine, service discovery engines such as the Semantic Discovery Service, and satisfiability engines including JSAT [SGPM04].

Recently, we have also expanded services to include explanation of text analytic platforms such as IBM's Unstructured Information Management Architecture and Inference web now not only explains answers from text analytic components (such as why Deb M is the same as Deborah M) but it also can point back to raw unstructured sources (as well as any structured sources) used in the derivation.

C 2.4 Common Logic (Hayes, IHMC)

Common Logic (CL, formerly called SCL) is a project, now close to completion, to design a standard logical notation with a straightforward model-theoretic semantics, suitable as a Semantic Web 'lingua franca' into which all other Semantic Web notations can be embedded without loss of meaning. [CL04]. The original inspiration of CL was as a modern successor to KIF [KIF95], and CL core syntax closely resembles KIF 3.0, but CL also has an XML syntax and has been modified and generalized in many respects to make it more suitable for Web use (URIs, dataypes , embedded basic types). It is also much more flexible in the constructions it allows (higher-order quantification, sequence quantification, role-value syntax, recursive axiom schemas) and has a thoroughly investigated model theory. One unique feature of CL is that a name may be used in multiple roles (individual, variable, function, relation, type or class name) freely, without damaging either the syntax or the meaning: this removes the need to check for mutual agreement on vocabulary conventions when combining knowledge from disparate sources. All the current standard semantic web languages (RDF, RDFS, DAML+OIL, OWL, SWRL) and most monotonic LP and 'rule' languages (DATALOG, RT) can be straightforwardly transcribed into CL without change in meaning [H05], and CL consistency can be checked – albeit with no performance guarantees - using a conventional first-order inference engine such as JTP (http://www.ksl.stanford.edu/software/JTP/) or VAMPIRE (http://www.cs.man.ac.uk/~riazanoa/Vampire/). CL is currently a candidate for ISO adoption as a logic standard [CL04]

C 2.5 Social Mechanisms for Establishing Trust (Feltovich, IHMC)

Elsewhere, we have attempted to encourage an expansion of thinking about the sources, nature, and diversity of mechanisms for establishing trust and coordination when people supported by computing systems are engaged in consequential work [FBJ+03]. One impetus for this direction has been a desire to make such networked systems acceptable to people by understanding what characteristics make a people and systems seem trustworthy (and actually be trustable) in their participation in important affairs, and just as importantly, to ensure, as in human societies, a kind of predictability.

Increasingly open and spontaneously emergent systems place special demands on the establishment and maintenance of trust among the parties who interact with these systems in significant ways. For such transactions to be trustworthy we have found in our research that transactions must conform to a pervasive system of diverse regulatory devices, from, for example, basic cultural "codes" of appropriate behavior and etiquette, to societal and organizational norms, to formal systems of law, for instance, tort law, contract law, and statutes related to the right of privacy. These complex systems of regulation, at many different levels and degrees of specificity, are fundamental to the establishment of order and predictability required for trust [FBJ04]. Adherence becomes more problematic as Web services increasingly become more spontaneously constructed and more loosely and locally controlled. We anticipate that a study of the relationship of cultural and societal mechanisms for regulation to more formal policy-based mechanisms

will be a fruitful source of new approaches and theory; in particular, it will inform the effort to produce adequate explanations for human use. Explanations may, for example, appeal to the authority of the information source; but such appeals may require sensitivity to the recipient's *perception* of the source as a social agent.

C 2.6 Provenance reasoning

Provenance services [GLM04, SM03] are a new approach to tracing how a given system—likely a distributed system composed of multiple services—has arrived at a particular result. *Execution provenance support* components record data passed among services in the generation of the result, allowing users to audit data transformations occurring at each step of the process. Various techniques for mutual authentication and non-repudiation can ensure the integrity of recording of provenance data. Besides its use for auditing, this data can be used to determine if previous results are still valid, if tools have changed, and so forth. *Service provenance support* components record data about individual service components to allow service providers to observe patterns leading to improvement of the service or for clients to select among multiple instances of a service based on historical data.

Currently provenance services depend on the cooperation of sources for provenance information that are associated with the generation process and tightly scripted. Mechanisms for representing and reasoning about provenance are currently quite simple, confining themselves to answering questions about topics such as what might have changed between different executions of the system. More powerful semantics and reasoning methods are needed to fully exploit provenance data.

C 3. Goals and methods

C 3.1 Integrating policy reasoning with trust negotiation

Bringing the high-level ontology reasoning of KAoS, and the transaction-oriented policies of TrustBuilder and PeerTrust into a common unified framework is one major part of the technical work plan. Lars Olson, working with IHMC in summer 2004, has made an initial study of the problems arising in approaching an integration KAoS policy ontology reasoning with the TrustBuilder ticket/handler architecture. This used an implementation of TrustBuilder as a web service handler, with a view to retrofitting to existing client-server code with minimal changes. This has revealed places where work is needed most immediately. TrustBuilder needs to be able to deal with policy changes dynamically (at present, these are loaded at startup); KAoS needs to be able to handle more complex queries against its action policies, for example to identify which agents are authorized to perform a given action, by tracing bindings to query variables. In addition, the RSA algorithm for signature checking is slow, and the system needs a performance and vulnerability analysis. Olson will be considering these issues in preparation for his doctoral thesis work at Illinois.

The BYU team are investigating the possibility of using XACML 2.0 as a basic policy language for an extended implementation of TrustBuilder. We will investigate how trust negotiation can be integrated into XACML There are several ways in which this integration might occur. First, the policy information point (in XACML, the 'place' where policies are enforced) accepts subject attribute information. In an open system, this could be information that was obtained during a trust negotiation. This information obtained during authentication could naturally be input into the XACML access control mechanism. Second, XACML, or a closely related language, could be adopted as the policy language for conducting the trust negotiation itself. Whether or not this is feasible requires a thorough analysis of the capabilities of XACML. We anticipate that it will require extensions to XACML. This part of the project will produce a set of recommendations for future extensions to XACML.

(There is a slight pun here on 'policy', which has a much more general sense in KAoS than in XACML, where it is restricted to matters of access control to information sources. Since the chief aim here is to integrate trust negotiation into a larger policy framework, we expect that XACML constructions will

be generally described in KAoS as instances of more general concepts. Some convergences are evident, e.g. between the KAoS notion of a 'policy guard' and the XACML notion of a 'point'.)

KAoS can describe policies which are much more general, and cover a wider range of agent actions, (resource-dependent, resource-limited, temporally restricted) policy types (permissions as well as prohibitions; overrides from higher authorities) and agent types than are required for the current TrustBuilder machinery, and it uses ontologies more centrally to describe all aspects of the domain; TrustBuilder limits its use of ontologies (so far) to taxonomies of transaction types. TrustBuilder, on the other hand, has connections to effective mechanisms for deployment, a thorough integration with secure transmission machinery, and a detailed acquaintance with concrete issues of information leakage and sensitive negotiation patterns. We expect each to benefit from a closer integration with the other. Some benefits of ontology use in trust negotiation have already been noted, particularly using ontologically expressed classifications to support more sophisticated negotiation strategies. For example, a negotiation server armed with an ontology which classifies a Cisco employee identification as an instance of the class 'enterprise issued employee identifications' can avoid revealing the existence of a special Cisco policy by casting a request for an identification in the more general form. In general, adding more knowledge to a negotiation can permit the negotiation to be more efficient and secure. Context information includes any information about the negotiation (what is trying to be accomplished, what kind of transaction, etc.). For instance, when applying for a loan, there is no need to disclose a medical credential if it were requested, even if the other party is able to satisfy the policy governing its disclosure. This can help thwart phishing attacks, as well as address the information leakage problem during trust negotiation. It can also permit more fine-grained control over the disclosure of freely available credentials. The ability for KAoS to reason and perform consistency checks over a wide range of actions and policy attributes provides wide scope for experimentation with contextual negotiations of this kind, referring if necessary to external databases maintained by other trusted authorities.

The initial mechanism we will use to link KAoS to TrustBuilder and XACML is by adding 'generic enforcers' to the TrustBuilder enforcer chain, which will allow KAoS reasoners to check credentials against more sophisticated policy constraints. This simple approach retains the transparency of the client/server interaction, but may require performance analysis in a realistically deployed system. This approach has been used to define enforcers that intercept SOAP messages from the CMU Semantic Matchmaker and filter results consistent with KAoS coalition policies. In a recent CoSAR-TS demonstration, these policies prevent the use of Gaoan resources [UBJ04b]. Recently IHMC has finished a first implementation of SOAP-enabled enforcer to understand arbitrary Semantic Web Service invocations so it can apply appropriate authorization policies to them. Additionally, it is equipped with a mechanism to perform obligation policies, which will be in the form of other Web Service invocations. For instance, some policy may require consultation or registration of performed transactions in some logging service available as a Web Service audit entity.

Opening up enforcement of policies to Web-based information raises issues of security and trust for that information itself, particularly when one considers that reasoners may be utilizing information from a variety of sources on the open Web network. We expect to make use of the emerging conventions for securing trustworthiness of Web information by the use of named warrant graphs [CBHS05] when considering secure ontology reasoning; part of the purpose of these conventions is to provide a secure provenance path from any asserted information on the Web to a digitally signed warrant which identifies the agency responsible for the assertion. This model can be generalized to handling assertions of policies which can be openly published but still be secure against misinterpretation. [BO04]

C 3.2 Trust Inference Catalog: a pragmatic foundational approach to Policy reasoning

All the current and proposed formalizations of trust reasoning use formal notations of limited expressivity, designed with the primary intention of supporting rapid run-time reasoning. KAoS uses OWL [OWL04], currently the most expressive standard Semantic Web formalism, which is essentially an RDF/XML transcription of the classical description logic *ALCQHIR*+; Rei [KFJ04] uses a rule language

with negation-as-failure and Horn expressivity; PeerTrust [NOW04] uses guarded Prolog (the distributed nature of the PeerTrust algorithm is irrelevant here.) All of these underlying formalisms have problems, when considered as a basic notation for expressing proofs, particularly those arising in trust negotiation and policy application.

OWL is simply not expressive enough. The expressive limitations of OWL have been noted in the KAoS project (it cannot express the class of *actions in which an agent modifies a resource owned by that agent*, for example) and elsewhere, notably in the OWL-Services project [MBH04]. Indeed, the expressive limitations of OWL are so severe that an entire W3C working group, the Semantic Web Best Practices Working Group, has been set up to explain how to use work-arounds.

The problems with Prolog-style notations are different, but we think more compelling for a notation in which to express trust reasoning: they are *logically invalid* (nonmonotonic). The problem here is that nonmonotonic strategies such as negation-as-failure, the unique name assumptionand default reasoning are essentially *enthymemes*: they omit 'hidden' assumptions which are necessary in order to firmly establish the conclusion. This often produces greatly increased efficiency at run time and simplifies the syntactic form of rules, but it is inherently dangerous. In describing delicate inferences, or inferences involving exceptional conditions, it is important to make the delicate assumptions (or exceptions) explicit in a checkable proof which will be the foundation for an intelligible explanation. For example, the use of negation-as-failure is essentially an unstated claim that the knowledge base is complete in some way, so that a failure to prove a sentence can be taken as a proof of its negation. A later discovery that some data was missing simply causes a Prolog-style reasoner to behave differently at run time, but is a *contradiction* when this assumption is made explicit; a fact which is likely to be critical in explaining the reason for drawing the false conclusion. Again, a 'normal' default is in fact an unstated assumption that something is in a certain category of normality, and exceptions from these norms must be noted explicitly in a fully correct derivation, again supplying a critical datum for generating an adequate explanation.

Seamons and Winslett have previously noted the importance of monotonicity in secure trust and policy reasoning in their NSF-funded joint work on policy reasoning.

A further limitation of many of these notations is an inability to make clear distinctions between use and mention of expressions, or between assertions and meta-assertions. These are often critically important in trust reasoning, however, as illustrated by the attention paid to query *forms* (as opposed to content) in the 'disguised transaction policy' work in TrustBuilder, described above. To fully expose the reasoning behind such policies represents a challenge for even the most expressive formal logic. Our approach will be to define the necessary epistemic concepts in explicit ontologies – logical theories - rather than modify the basic CL notation; experience with fixing the semantics of a fully expressive logic strongly suggests that changing the logic itself should be a last resort, one we do not expect to take during the course of this project.

We will track the KAoS-TrustBuilder integration, transcribing the reasoning into Common Logic and registering any special inference patterns used as derivation rules in Proof Markup Language (or its extensions) and incorporated into Inference Web. This will be done partly manually and, where possible, semi-automatically (eg from OWL). One goal of this is to make all assumptions completely explicit in the statement of the inference rules; we expect that this will require a concomitant effort to write coherent common logic ontologies for aspects of the trust domain, to help capture the required inferences: these ontologies are part of the intended end-product.

The point here is not, in the first instance, to replace the efficient OWL and rule-based inference engines by a logical theorem-prover, but rather to create a logically coherent audit trail of the reasoning processes as a machine-checkable proof which can be used to generate explanations, using the Inference Web technology. There is also a more scientific interest to this aspect of the project, which is to locate, describe and study the precise kinds of reasoning steps taken, and thereby formalize the reasoning involved. As noted, much of it is logically challenging, particularly the epistemic assumptions about information leakage. In some cases the best we can do may be to register a rule using text strings as descriptors of the rule conditions; this will weaken the explanation generation process at that point, but not totally disable it.

C 3.3 Generating trust and policy explanations

For this effort, we will leverage the Inference Web infrastructure to improve trust in answers by exposing sources, meta-information, and information manipulations that were applied to obtain answers. We will use the Proof Markup Language as a starting point, which in turn uses the Ontology Web Language to represent proof information. We will work with a requirements-driven and use-case-driven methodology to either identify that the Proof Markup Language is adequate in its current form or identify necessary extensions to encode justifications of answers.

We will also broaden our work on tactics used to abstract and summarize proofs and turn them into more understandable explanations. We will expand upon our work on trust networks [ZPM05] to include an explanation capability that exposes trust levels of sources and answers.

The impact of providing access to meta information about sources used in question answering is that users can now be provided with a quick summary of the recency and reliability of sources used to generate answers. They may also have access to the assumptions on which answers are based. When text analytic components are used, they may also have access to the raw sources that were used to obtain an answer. The impact of providing access to the derivation path is that users may be confident that they can ask follow-up questions at whatever level of granularity is necessary to convince themselves that the information manipulations were appropriate and relevant. Simultaneously, they will be provided with an interactive browser view of the answer and the information that led a system to obtaining that answer. The viewer is embeddable in other systems so that end users may customize views according to their needs.

C 3.4 Discovering a tractable trust logic

This inference catalog also has another, more exploratory, purpose, as input to an effort to characterize a useful optimized reasoning engine for trust and policy reasoning. All current proposals for such engines take some existing efficient subcase off the shelf, as it were. Even when such reasoners work well in practice, there is no particular reason why they should. Efficiency in a reasoner arises from bounding the expressiveness of the logic. The computational intractability of expressive logical languages arises from the assumption that a reasoner for a language should be able to accept, and draw conclusions from, *any* syntactically legal expression of the language. As has often been observed, however, in practice very few of the syntactically legal expression forms of any recursive language actually arise in real applications. Theoretical worst-case complexity analyses frequently underestimate the practical utility of inference methods largely because they are bounded by theoretically possible cases which never arise in practice. Part of our purpose, then, is to identify and catalog the forms of knowledge that are in fact used, with a view to *discovering* the appropriate subset of a fully expressive logic which is actually needed for practical trust reasoning. One can view this as a process of inventing or discovering a 'logic for trust', except that, crucially, there is no expectation that the resulting set of expressions must form a language with a recursive syntax. For example, it might well be bounded by restrictions on the syntactic depth of nesting of certain constructions, or by branching factors in syntactic graphs.

We expect to be to locate a set of expressions – patterns might be more accurate – which is adequately expressive along all the necessary dimensions for trust and policy reasoning, but with tractable inference characteristics. In part, this optimism is based on recent work identifying new tractable subcases of full first-order logic such as the guarded fragment [ABN98] and its variations. These strongly suggest that the characteristic model-theoretic properties which signal robust decideability in theory, and reasonable practical tractability in practice, are found in a wider class of subcases than had formerly been expected [H02]. In particular, these advantageous computational properties are not restricted to description logics.

C 3.5 Proofs and Provenance

The notion of provenance can be generalized to be closer to that of proof, where provenance is seen as any information that can be used to trace back to the source of a justification, so that provenance can be

inferred by trusted methods from general historical information from various sources. The generality of this approach introduces potential complexities and requires close attention to issues such as reliability of the provenance reasoner and the notion of meta-provenance.

In collaboration with IHMC collaborator Luc Moreau at University of Southampton (who is funded through European grants), we will apply and evaluate our research in the context of the Southampton provenance architecture. A letter of support outlining our planned collaboration can be found in section J. We anticipate three major topics of investigation: the relationship between provenance and proofs in the context of Semantic Web inferences; reasoning over provenance in order to infer trust; and applying policies to provenance architectures in order to enforce their safety and liveness. In collaboration with Moreau and his colleagues, we anticipate the evaluation of our results in the context of scenarios in the following domains: e-commerce with IBM, aerospace engineering with German Aerospace, particle physics with CERN, computational steering of physical science models with the UK RealityGrid project, and medical decision support with the University Polytecnica de Catalunya in Barcelona.

C 3.6 Understanding the Social Mechanisms for Establishing Trust and Coordination

We will draw on experience in legal, cultural, and social contexts of policy to address issues arising from the deployment of these mechanisms in real-world organizations and systems, with particular emphasis, in this project, on the perception of trustworthiness in Web-based transactions. As noted, this seems to depend on the perception that the system is applying and conforming to understood societal and organizational norms or the system's ability to explain and justify how these regulatory systems have been applied, reconsidered, prioritized, deconflicted, etc. These include legal codes and administrative boundaries and policies, understood by professional users, but they also include broader social "codes" of fairness, appropriateness to the circumstances and perceived rights of privacy and obligations of accountability. We will apply these modes of social and psychological analysis to analyze the longitudinal perception of system behavior among its human users and observers, and relate these observations to the formal policy descriptions in use. We will explore and describe the nature of explanation that engenders or impedes trust.

C 3.7 Integration, Validation and Evaluation Plan

In addition to cooperative links with team members already described, we will test these ideas as a team by focusing on the design of a information-handling system for the Emergency Operations Center (EOC) of the City of Champaign, Illinois. As in all US cities, Champaign's EOC is activated during natural, technological, and human-caused disasters, to serve as a central point of coordination and control. Potential causes of EOC activation range from tornadoes and straight-line winds to hazardous material spills, hostage situations, or random shooting incidents. Co-PI Winslett has worked with City officials (city manager, city IT director, deputy fire chief, and director of the Champaign County GIS Consortium) to identify the information technology that the City would most like to have added to its current EOC. The desired technology is the ability to superimpose real-time data feeds on top of an interactive GIS product that shows Champaign buildings and roads (see attached letters of support). The interactive GIS product will be an enhanced version of products already produced by the Champaign County GIS Consortium. The data feeds will come from 911 call databases and from a variety of sensor sources, including chemical sensors in storm sewers and on street light poles, temperature sensors on roads, wind sensors on light poles, and cameras located around the city. These data feed sources are owned by a wide variety of organizations that do not ordinarily share their data directly with the City of Champaign: the city of Urbana, University of Illinois, Champaign county (population 180,000), federal and state highway authorities, local schools and hospitals, METCAD (the 911 call center), and individual companies. The major players from these government organizations and from the first responder agencies all know one another personally and have a level of mutual respect that means that social factors do not pose a major barrier to allowing increased information sharing during a disaster.

The purpose of the Champaign testbed is to provide a focused, realistic example on which to test our ideas and integrate our approaches. In particular, as the cornerstone of the authorization architecture, we will deploy a stand-alone version of TrustBuilder that supports legacy applications, such as 911 and sensor databases and camera feeds. This version of TrustBuilder will be in charge of enforcing the access control policies agreed upon by the EOC stakeholders and data providers, as expressed in KAoS policies. In the event of a declared emergency or official drill, these policies will allow a software agent acting on behalf of an authorized EOC stakeholder to obtain a token (e.g., Kerberos ticket or login/password) that can be used to obtain information from a legacy application, without changing the application itself. If access is denied by TrustBuilder, an Inference Web system will explain the rationale for the denial. All access attempts and the resulting authorization decisions will be logged using tamper-evident LogCrypt [Jason Holt from BYU] logging facilities, to serve as an audit trail of policy-related events and inferences. The resulting prototype is intended to serve as a demonstration of concept and a potential springboard for Champaign (and potentially hundreds of other mid-size cities) to seek funding for a fully operational system.

The interest of this application area, from the point of view of this project, lies in the range of information types, sources, stake holders and owners, leading to semantic integration issues and a wide variety of authorization policies for information access; the need to apply unusual, perhaps conflicting, policies concerning information access and communication priorities in emergency situations; the various complexities of balancing normal privacy concerns against the access needs of the emergency personnel; and in the need for rapid and comprehensible explanations and/or displays of information concerning policies, policy application decisions and data access to the various agencies and controllers involved, both during the critical phase of emergency activity and after the event, in post-disaster analysis for training, reporting and public relations.

C 4. Management Plan

Research Teams and Specific Roles of the Investigators. In this collaborative proposal, the four institutions will organize their activities in research teams. PI Jeff Bradshaw (IHMC) will coordinate the research activity as a whole. Each team will have a research coordinator, a PI from one of the participating institutions. The research coordinator will inform the others of the research milestones achieved, and will ensure that the research activities are executed in a timely manner and that they constitute a coherent whole. The project as a whole will have Web sites that are mirrored in each of the participating institutions. In addition, each research coordinator will maintain the part of the Web site that he or she coordinates to make sure that reports, tools, and announcements of interest to the research activity are posted and maintained.

- **Policy Inference Catalog**. Co-PI Pat Hayes (IHMC) will coordinate this research activity.
- **Explanation**. Linked co-PI Deborah McGuinness (Stanford KSL) will coordinate this research activity.
- **Trust Negotiation and XACML Integration**. Linked PI Kent Seamons (BYU) will coordinate this research activity.
- **KAoS and TrustBuilder Integration**. PI Jeff Bradshaw (IHMC) will coordinate this research activity.
- **Emergency Response Testbed.**. Linked PI Marianne Winslett (Illinois) will coordinate this research activity.
- **Social Mechanisms.** Paul Feltovich (IHMC) will coordinate this research activity.

Although these define areas of responsibility and coordination, each of these activities involves cooperation with people at other institutions, and we will use established techniques of remote collaboration, including group teleconferencing, email, IRC and instant-messaging communication, to keep in touch, as well as direct visits. An essential component of the coordination plan is an annual workshop. IHMC, BYU, Illinois, and Stanford will host these workshops alternately in each year. All investigators, along with their graduate students, will be expected to participate in the workshops. At the first workshop, it is expected that the theoretical foundations will be sufficiently developed to allow for discussion and presentation. Such a venue will permit further refinements and enhancements, and also serve as an education forum.

Although this entire team as a whole is new, many existing collaborations between subsets of the team have already supported extended research programs; Winslett and Seamons have worked productively together for five years; Hayes and McGuinness have collaborated on many projects, including PML, and are colleagues on several W3C working groups. Bradshaw, Hayes and McGuinness are all active in the DARPA DAML effort, which has continued for four years. Bradshaw and Feltovich have collaborated extensively on analyses of human-agent teamwork, coordination and trust through policy use. Students from BYU have worked on summer projects at IHMC: most notably, Lars Olson went from a masters at BYU to an internship at IHMC with Bradshaw and is now a doctoral student at Illinois under Winslett. We will continue to collaborate in these useful ways, particularly by lending students to other institutions for extended work and study experiences.

Management of the Project Activities in Each Institution. The respective project activities will be managed by the investigators at their respective institutions. The project is multidisciplinary in nature and the project-related workshops, conference presentations, and publications will reflect this. Moreover, frequent posting on the our project web site will enhance visibility and accessibility of the project outcomes across the disciplines and to the public.

Evaluation Plan. We will test the algorithms extensively in our research laboratories. To foster collaboration and integration of research results among the team members, we will jointly provide case studies on which we can test the language and algorithms. The chief evaluation of the project will be the Champaign Emergency response testbed effort coordinated by Winslett.

C 5. Results From Prior NSF Support

(Note, this lists priors of all PIs on all the linked cooperating projects)

ITR: Responding to the Unexpected, IIS-0331707 & 0331690, $12,500,000, Oct. 2003-Sep. 2008. S. Mehrotra, UC-Irvine, PI; Co-PIs include M. Winslett. The RESCUE project is focusing on disaster management in the LA area (fires, earthquakes, floods, terrorism, etc.), with cooperation from LA-area police, fire, transportation, and government officials. The project is working to radically transform the ability of responding organizations to gather, manage, use and disseminate information both within emergency response networks and to the general public. Site http://www.itr-rescue.org gives full details.

ITR: Automated Trust Negotiation in Open Systems, CCR-0325951, $1,750,000, Oct. 2003-Sep. 2008, K. Seamons, PI; Co-PIs include M. Winslett. This project complements ours by proposing the development of needed extensions to the *RT* authorization policy language family and a lightweight compliance checker for *RT*; new trust negotiation message conventions for interoperability and negotiation strategies compatible with the *RT* extensions; extensions to the TrustBuilder prototype supporting an interface for access control frameworks such as GAA-API.

C 5.1 Related Work funded from other sources

The DARPA CoABS-sponsored Coalition Operations Experiment (CoAX) (http:// www.aiai.ed.ac.uk /project/ coax/) [ABB+02, ABK+03] was an international cooperation spanning four countries and twenty-eight partners. It modeled military coalition operations and implemented agent-based systems to mirror coalition structures, policies, and doctrines. CoAX aimed to show that the agent-based computing paradigm offers a promising new approach to dealing with issues such as the interoperability of new and legacy systems, the implicit nature of coalition policies, security, and recovery from attack, system failure, or service withdrawal. The most recent CoAX-related work, sponsored by the DARPA DAML program, also investigated issues in composition of semantic web services consistent with negotiated policy constraints [UBJ+04a]. KAoS provided mechanisms for overall management of coalition organizational structures represented as domains and operational constraints represented as policies [BUJ+03], while *Nomads* enabled strong mobility, resource management, protection from denial-of-service attacks for untrusted

agents that run in its environment, and transparent filtering and transformation of data feeds from sensors SCB03].

Within the DARPA Ultra*Log program (http://www.ultralog.net) we collaborated with CougaarSoft to extend and apply KAoS policy and domain services to assure the scalability, robustness, and survivability of logistics functionality in the face of information warfare attacks or severely constrained or compromised computing and network resources. In agent societies of over a thousand agents and hundreds of policies, dynamic policy updates can be committed, deconflicted, and distributed across multiple hosts in a matter of seconds, and responses to policy authorization queries average less than 1 ms.

As part of the Army Research Lab Advanced Decision Architectures Consortium, we have been investigating the use of KAoS and *Nomads* technologies to enable soldiers in the field to use agents from handheld devices to perform tasks such as dynamically tasking sensors and customizing information retrieval [SBC+03; SCB+03]. The combination of FlexFeed's implementation of agile computing and KAoS policies provide a technical foundation for this work [SBC+03a]. The approach was validated through our participation in MOUT exercises at Ft. Benning in 2003 and 2004.

An application focused more on the social aspects of agent policy is within the NASA Cross-Enterprise and Intelligent Systems Programs. Here we are investigating the integration of Brahms, an agent-based design toolkit that can be used to model and simulate realistic work situations in space [CSS98; S01], with KAoS policy-based teamwork models and Nomads's strong mobility and resource control capabilities for use in highly-interactive autonomous systems. The same approach is also being generalized for use by combinations of astronauts and mobile robots for planetary surface exploration [BAA+04; BSA+03]. Each year, we participate with members of the NASA Mobile Agents project as part of a two-week exercise with robots and people in the role of astronauts performing planetary surface exploration at the Mars Desert Research Station in southern Utah [SBA+03].

The Office of Naval Research (ONR) is supporting research to extend this work on effective human-agent interaction to unmanned vehicles and other autonomous systems that involve close, continuous interaction with people. As one part of this research IHMC and University of South Florida are developing a new robotic platform with carangiform (fish-like) locomotion, specialized robotic behaviors for humanitarian demining, human-agent teamwork, agile computing, and mixed-initiative human control. As part of this effort KAoS and FlexFeed have been integrated with TRIPS [BAA+04] for joint handling of mixed-initiative dialogue and adjustable autonomy issues, and the SFX hybrid robotic architecture (http://crasar.eng.usf.edu/research/publications.htm) to allow policy-based governance of selected robotic behaviors. To facilitate local field tests, we have established the IHMC Robot Ranch comprising a growing variety of ground-based and airborne robotic platforms.

We are investigating issues in adjustable autonomy and mixed-initiative behavior for software assistants in an office environment as part of the DARPA EPCA (SRI CALO) program's Multi-Modal Dialogue team [BJKT04]. Under funding from DARPA's Augmented Cognition Program, we are taking the challenge of effective human-agent interaction one step further as we investigate whether a general policy-based approach to the development of cognitive prostheses can be formulated, in which human-agent teaming could be so natural and transparent that robotic and software agents could appear to function as direct extensions of human cognitive, kinetic, and sensory capabilities [BBR+03].

We are also conducting research leading to a better understanding of the social mechanisms for establishing trust and coordination within advanced networked systems, with analogues to human and animal cultures [FBJ+04]. March and Simon made a well known analysis of organizational dynamics in terms of two basic logics of action: the *logic of consequence* and the *logic of appropriateness* [M89, M89a, MS93]. Our primary interest in the proposed research is to investigate how the loop between action and results can be closed, moving organizations from the logic of appropriateness to the logic of consequence. We would like people to be able to trace and understand the positive and negative effects of policies that have been put into force, to determine where these effects are coming from, and to discover how policies might be adjusted for greater effectiveness [22]. We have begun to apply this research as part of an

investigation of requirements for policy-based information access and analysis within intelligence applications.

D. References Cited

[ABB+02] Allsopp, D., Beautement, P., Bradshaw, J. M., Durfee, E., Kirton, M., Knoblock, C., Suri, N., Tate, A., & Thompson, C. (2002). Coalition Agents eXperiement (CoAX): Multi-agent cooperation in an international coalition setting. *A. Tate, J. Bradshaw, and M. Pechoucek (Eds.), Special issue of IEEE Intelligent Systems*, 17(3), 26-35.

[ABK+03] Allsopp, D., Beautement, P., Kirton, M., Tate, A., Bradshaw, J. M., Suri, N., & Burstein, M. H. (2003). The Coalition Agents Experiment: Network-Enabled Coalition Operations. *Journal of Defence Science*, 8(3), 130ff.

[ABN98] Andreka, H., van Bentham, J., & Nemeti, I. (1998). Modal languages and bounded fragments of predicate logic. *J. Phil. Logic*, 27(3), 217-274.

[BAA+04] Bradshaw, J. M., Acquisti, A., Allen, J., Breedy, M. R., Bunch, L., Chambers, N., Feltovich, P., Galescu, L., Goodrich, M. A., Jeffers, R., Johnson, M., Jung, H., Lott, J., Olsen Jr., D. R., Sierhuis, M., Suri, N., Taysom, W., Tonti, G., & Uszok, A. (2004). Teamwork-centered autonomy for extended human-agent interaction in space applications. *AAAI 2004 Spring Symposium*. Stanford University, CA, AAAI Press,

[BBR+03] Bradshaw, J. M., Beautement, P., Raj, A., Johnson, M., Kulkarni, S., & Suri, N. (2003). Making agents acceptable to people. In N. Zhong & J. Liu (Ed.), *Intelligent Technologies for Information Analysis: Advances in Agents, Data Mining, and Statistical Learning.* (pp. in press). Berlin: Springer Verlag.

[BCM+03] Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P. (2003). *The Description Logic Handbook*. Cambridge, England: Cambridge University Press.

[BJKT04] Bradshaw, J. M., Jung, H., Kulkarni, S., & Taysom, W. (2004). Dimensions of adjustable autonomy and mixed-initiative interaction. In M. Klusch, G. Weiss, & M. Rovatsos (Ed.), *Computational Autonomy.* (pp. in press). Berlin, Germany: Springer-Verlag.

[BO04] Bizer, C. & Oldakowski,R. Using Context- and Content-Based Trust Policies on the Semantic Web, *Poster at WWW2004* http://www.wiwiss.fu-berlin.de/suhl/bizer/SWTSGuide/p747-bizer.pdf

[BSA+03] Bradshaw, J. M., Sierhuis, M., Acquisti, A., Feltovich, P., Hoffman, R., Jeffers, R., Prescott, D., Suri, N., Uszok, A., & Van Hoof, R. (2003). Adjustable autonomy and human-agent teamwork in practice: An interim report on space applications. In H. Hexmoor, R. Falcone, & C. Castelfranchi (Ed.), *Agent Autonomy.* (pp. 243-280). Kluwer..

[BUJ+03] Bradshaw, J. M., Uszok, A., Jeffers, R., Suri, N., Hayes, P., Burstein, M. H., Acquisti, A., Benyo, B., Breedy, M. R., Carvalho, M., Diller, D., Johnson, M., Kulkarni, S., Lott, J., Sierhuis, M., & Van Hoof, R. (2003). Representation and reasoning for DAML-based policy and domain services in KAoS and Nomads. *Proceedings of the Autonomous Agents and Multi-Agent Systems Conference (AAMAS 2003)*. Melbourne, Australia, New York, NY: ACM Press,

[CBHS05] Carroll, J., Bizer, C., Hayes, P., Stickler, P. (2005) *Named Graphs, Provenance and Trust* (HP technical report HPL-2004-57 (2004) http://www.hpl.hp.com/techreports/2004/HPL-2004-57.pdf accepted for 14th International WWW conference, Keio, Japan 2005)

[CL04] Altheim, M., Delugash, H., Hayes, P., Menzel, C., Tamet, T., Anderson, W., & Sowa, J. *CL: A proposal for a Web standard logic (ISO Work Item 24707 under JTC 1/SC32;* http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39175&scopelist=PROGR

[S01] Sierhuis, M. (2001) *Brahms: A Multi-Agent Modeling and Simulation Language for Work System Analysis and Design*. Doctoral Thesis University of Amsterdam.

[CSS98] Clancey, W. J., Sachs, P., Sierhuis, M., & van Hoof, R. (1998). Brahms: Simulating practice for work systems design. *International Journal of Human-Computer Studies*, 49, 831-865.

[FBJ+04] Feltovich, P., Bradshaw, J. M., Jeffers, R., Suri, N., & Uszok, A. (2004). Social order and adaptability in animal and human cultures as an analogue for agent communities: Toward a policy-based approach. In *Engineering Societies in the Agents World IV*. Berlin, Germany: Springer-Verlag.

[FBJ+03] Feltovich, P., Bradshaw, J. M., Jeffers, R., & Uszok, A. (2003). Social order and adaptability in animal, human, and agent communities. *Proceedings of the Fourth International Workshop on Engineering Societies in the Agents World,* (pp. 73-85). Imperial College, London,

[FBU04] Feltovich, P., Bradshaw, J. M., & Uszok, A. (2004). *Moving from the logic of appropriateness to the logic of consequence: Leveraging operational feedback within policy-governed systems*. Institute for Human and Machine Cognition, March.

[GLM04] Groth, P., Luck, M., & Moreau, L. A protocol for recording provenance in service-oriented grids. (2004) In Proceedings of the 8th International Conference on Principles of Distributed Systems (OPODIS'04)< Grenoble, France, Dec 2004 http://www.ecs.soton.ac.uk/~lavm/papers/opodis04.pdf

[H05] Hayes, P. Translating Semantic Web languages into SCL. Preparing for submission, http://www.ihmc.us/users/phayes/CL/SW2SCL.html

[H02] Hladik, J. (2002). Implementation and optimization of a tableau algorithm for the guarded fragment. *Proceedings of the Federated Logic Conference (FLoC-02),* http://floc02.diku.dk/TABLEAUX/program.html). Copenhagen, Denmark,

[JCJ+03] Johnson, M., Chang, P., Jeffers, R., Bradshaw, J.M., Soo, V-W., Breedy M. R., Bunch, L., Kulkarni, S., Lott, J., Suri, N., & Uszok, A. (2003). KAoS semantic policy and domain services: An application of DAML to Web services-based grid architectures. In *Proceedings of the AAMAS 03 Workshop on Web Services and Agent-Based Engineering.* Melbourne, Australia, July. (To appear in a forthcoming volume from Kluwer.)

[KIF95] Genesereth, M. (ed) *Knowledge Interchange Format Specification* http://www.csee.umbc.edu/kse/kif/

[KFJ04] Kagal , L., Finin, T. & Joshi, A. A Policy Language for A Pervasive Computing Environment" In Collection, *IEEE 4th International Workshop on Policies for Distributed Systems and Networks*, June 2003. http://www.cs.umbc.edu/~finin//papers/policy03.pdf

[M89] March, J. G. (1989). *Decisions and Organizations*. Blackwell Publishers.

[M89a] March, J. G. (1989). *Rediscovering Institutions*. Free Press.

[MS93] March, J. G., & Simon, H. A. (1993). *Organizations*. (Second ed.), Cambridge, England: Blackwell Publishers.

[McP04a] Deborah L. McGuinness and Paulo Pinheiro da Silva. Explaining Answers from the Semantic Web: The Inference Web Approach. Web Semantics: Science, Services and Agents on the World Wide Web Special issueEdited by K.Sycara and J.Mylopoulis. Volume 1, Issue 4. October 2004. http://www.websemanticsjournal.org/ps/pub/2004-22

[McP03] Deborah L. McGuinness and Paulo Pinheiro da Silva. Infrastructure for Web Explanations. In Proceedings of 2nd International Semantic Web Conference (ISWC2003), Sanibel Is., FL, USA. Springer, October 2003.

[MBH+04] Martin, D., Burstein, M., Hobs, J., Lassila, O.,McDermott, D., Narayanan, S., Paolucci, M.,Parsia, B., Payne, T., Sirin, E., Srinivasan N. & Sycara, K. OWL-S: *Semantic Markup for Web Services* (W3C Member submission 22 November 2004)

[NOW04] Nejdl, W., Olmedilla, D., Winslett,M. PeerTrust: Automated Trust Negotiation for Peers on the Semantic Web. In *Proc. of the Workshop on Secure Data Management in a Connected World (SDM'04)* in conjunction with *30th International Conference on Very Large Data Bases, Aug.-Sep. 2004, Toronto, Canada* http://www.w3.org/Submission/2004/SUBM-OWL-S-20041122/

[OWL04] *Web Ontology Language*. W3C website http://www.w3.org/2004/OWL/

[PHH04] Patel-Schneider, P., Hayes, P., & Horrock, I. (2004). *OWL Web Ontology Language: Semantics and Abstract Syntax* (W3C Recommendation), http://www.w3.org/TR/2004/REC-owl-semantics-20040210/

[PMF05] Paulo Pinheiro da Silva, Deborah L. McGuinness and Richard Fikes**. A** Proof Markup Language for Semantic Web Services. *Information Systems*. To appear. Also *Technical Report KSL-04-01*, Knowledge Systems Laboratory, Stanford University, USA, 2004.

[PMM03] Paulo Pinheiro da Silva, Deborah L. McGuinness, Rob McCool, Knowledge Provenance Infrastructure. IEEE Data Engineering Bulletin Vol.26 No.4, pages 26-32, December 2003. www.ksl.stanford.edu/people/dlm/papers/provenance-abstract.html

[S02] Sierhuis, M. (2002). *Brahms - Modeling and Simulating Work Practice*. Univ. of Amsterdam Press,

[SBA+03] Sierhuis, M., Bradshaw, J. M., Acquisti, A., Van Hoof, R., Jeffers, R., & Uszok, A. (2003). Human-agent teamwork and adjustable autonomy in practice. *Proceedings of the Seventh International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*. Nara, Japan,

[SBC+03] Suri, N., Bradshaw, J. M., Carvalho, M., Breedy, M. R., Cowin, T. B., Saavendra, R., & Kulkarni, S. (2003). Applying agile computing to support efficient and policy-controlled sensor

information feeds in the Army Future Combat Systems environment. *Proceedings of the Annual U.S. Army Collaborative Technology Alliance (CTA) Symposium*.

[SBC+03a] Suri, N., Bradshaw, J. M., Carvalho, M., Cowin, T. B., Breedy, M. R., Groth, P. T., & Saavendra, R. (2003). Agile computing: Bridging the gap between grid computing and ad-hoc peer-to-peer resource sharing. O. F. Rana (Ed.), *Proceedings of the Third International Workshop on Agent-Based Cluster and Grid Computing*. Tokyo, Japan,

[SCB+03] Suri, N., Carvalho, M., Bradshaw, J. M., Breedy, M. R., Cowin, T. B., Groth, P. T., Saavendra, R., & Uszok, A. (2003). Mobile code for policy enforcement. *Policy 2003*. Como, Italy,

[SGPM04] Pavel Shvaiko, Fausto Giunchiglia, Paulo Pinheiro da Silva and Deborah L. McGuinness. Web Explanations for Semantic Heterogeneity Discovery. *Technical Report KSL-04-02*, Knowledge Systems Laboratory, Stanford University, USA, 2004. http://www.ksl.stanford.edu/KSL_Abstracts/KSL-04-02.html

[SM03] Szomszor, M., & Moreau, L. (2003). Recording and reasoning over data provenance in web and grid services. *Proceedings of the International Conference on Ontologies, Databases, and Applications of Semantics (ODBASE 2003). Volume 2888 of the Lecture Notes in Computer Science Series,* (pp. 603-620). Catania, Sicily, Italy, Springer-Verlag,

[TBJ+03] Tonti, G., Bradshaw, J. M., Jeffers, R., Montanari, R., Suri, N., & Uszok, A. (2003). Semantic Web languages for policy representation and reasoning: A comparison of KAoS, Rei, and Ponder. In D. Fensel, K. Sycara, & J. Mylopoulos (Ed.), *The Semantic Web—ISWC 2003. Proceedings of the Second International Semantic Web Conference, Sanibel Island, Florida, USA, October 2003, LNCS 2870.* (pp. 419-437). Berlin: Springer.

[TMB+04] Tonti, G., Montanari, R., Bradshaw, J. M., Bunch, L., Jeffers, R., Suri, N., & Uszok, A. (2004). Automated Generation of Enforcement Mechanisms for Semantically-rich Security Policies in Java-based Multi-Agent Systems. *Proc. First IEEE Symposium on Multi-Agent Security and Survivability* (Mass2004)

[UBJ+04] Uszok, A., Bradshaw, J. M., Jeffers, R., Tate, A., & Dalton, J. (2004) Applying KAoS Services to Ensure Policy Compliance for Semantic Web Services Workflow Composition and Enactment. In S.A. McIlraith et al. (Eds.): *Proc. ISWC 2004*, LNCS 3298, pp. 425-440. Berlin: Springer-Verlag

[UBJ+04a] Uszok, A., Bradshaw, J. M., Jeffers, R., Johnson, M., Tate, A., Dalton, J., & Aitken, S. (2004). Policy and contract management for semantic web services. *AAAI 2004 Spring Symposium Workshop on Knowledge Representation and Ontology for Autonomous Systems*. Stanford University, CA, AAAI Press,

[UBJ04b] Uszok, A., Bradshaw, J., Jeffers, R., Johnson, M., Tate A., Dalton, J., Aitken, S. (2004). KAoS Policy Management for Semantic Web Services. In IEEE Intelligent Systems, Vol. 19, No. 4, July/August, p. 32-41.

[ZPM05] Ilya Zaihrayeu, Paulo Pinheiro da Silva and Deborah L. McGuinness. IWTrust: Improving User Trust in Answers from the Web. *Proceedings of 3rd International Conference on Trust Management (iTrust2005),* Springer, Rocquencourt, France, 2005.

Jeffrey M. Bradshaw, PhD jbradshaw@ihmc.us

a. Professional Preparation

University of Utah	Psychology	BA	1979
Brigham Young University	Experimental Psychology		1980
University of Washington	Cognitive Science	Ph.D.	1996

b. Appointments

2000- Senior Research Scientist, Institute for Human and Machine Cognition, Pensacola, FL (Kenneth Ford)

1985-2000 Associate Technical Fellow, Principal Investigator, Intelligent Agent Technology Research and Technology, Boeing Information and Support Services, Seattle, Washington (Ken Neves)

1993-1999 Co-Principal Investigator, Post-transplant Support Technology Project Fred Hutchinson Cancer Research Center, Seattle, Washington (Keith Sullivan)

1993-1994 Principal Investigator, European Institute of Cognitive Sciences and Engineering (EURISCO), Toulouse, France (Guy Boy)

c. Related Publications

(i) Closely Related Publications

Bradshaw, J. M., Jung, H., Kulkarni, S., Johnson, M., Feltovich, P., Allen, J., Bunch, L., Chambers, N., Galescu, L., Jeffers, R., Suri, N., Taysom, W., & Uszok, A. (2005). Toward trustworthy adjustable autonomy in KAoS. In R. Falcone *et al.* (Eds.), *Trusting Agents for Trustworthy Electronic Societies.* LNAI. Berlin: Springer, in press.

Bradshaw, J.M., Jung, H., Kulkarni, S., & Taysom, W. (2004). Dimensions of adjustable autonomy and mixed-initiative interaction. In M. Nickles, M. Rovatsos & G. Weiss (Eds.), *Agents and Computational Autonomy: Potential, Risks, and Solutions*. Lecture Notes in Computer Science, Vol. 2969. Berlin: Springer-Verlag, pp. 17-39.

Bradshaw, J. M., Beautement, P., Breedy, M., Bunch, L., Drakunov, S. V., Feltovich, P. J., Hoffman, R. R., Jeffers, R., Johnson, M., Kulkarni, S., Lott, J., Raj, A., Suri, N., & Uszok, A. (2004). Making agents acceptable to people. In N. Zhong and J. Liu (Eds.), *Intelligent Technologies for Information Analysis: Advances in Agents, Data Mining, and Statistical Learning,* Berlin: Springer Verlag, pp, 361-400.

Bradshaw, J. M., Uszok, A., Jeffers, R., Suri, N., Hayes, P., Burstein, M. H., Acquisti, A., Benyo, B., Breedy, M. R., Carvalho, M., Diller, D., Johnson, M., Kulkarni, S., Lott, J., Sierhuis, M., & Van Hoof, R. (2003). Representation and reasoning for DAML-based policy and domain services in KAoS and Nomads. *Proceedings of the Autonomous Agents and Multi-Agent Systems Conference (AAMAS 2003).* 14-18 July, Melbourne, Australia. New York, NY: ACM Press, pp. 835-842.

Bradshaw, J.M., Suri, N., Canas, A.J., Davis, R., Ford, K.M., Hoffman, R., Jeffers, R., and Reichherzer, T. Terraforming Cyberspace. *IEEE Computer* (July 2001), pp 48-56

(ii) Other Significant Publications

Bradshaw, J.M. (Ed.) (1997). *Software Agents.* Cambridge, MA: AAAI/MIT Press.

Bradshaw, J. M., Greaves, M., Holmback, H., Jansen, W., Karygiannis, T., Silverman, B., Suri, N., & Wong, A. (1999). Agents for the masses? In J. Hendler (Ed.) Special issue on agent technology, *IEEE Intelligent Systems*, March/April, 53-63.

Bradshaw, J. M., Cabri, Giacomo & Montanari, Rebecca (2003). Taking back cyberspace. *IEEE Computer,* July, pp. 89-92.

Bradshaw, J. M., Sierhuis, M., Acquisti, A., Feltovich, P., Hoffman, R., Jeffers, R., Prescott, D., Suri, N., Uszok, A., & Van Hoof, R. (2003). Adjustable autonomy and human-agent teamwork in practice: An interim report on space applications. In H. Hexmoor, R. Falcone, & C. Castelfranchi (Ed.), *Agent Autonomy*. Dordrecht, The Netherlands: Kluwer, pp. 243-280. Republished in RTO-MP-088 (ISBN: 92-837-0031-7). NATO Research and Technology Organization Report, pp. KN6 1-30.

Uszok, A., Bradshaw, J. M., Johnson, M., Jeffers, R., Tate, A., Dalton, J., & Aitken, S. (2004). KAoS policy management for semantic web services. *IEEE Intelligent Systems,* July/August, 19(4), pp. 32-41.

d. Synergistic Activities

2004 Member, Editorial Board, International Journal of Human-Computer Studies (IJHCS)

2004 Program Chair, First IEEE Symposium on Multi-Agent Security and Survivability (MASS 2004), Drexel Univ, 30-31 Aug

2002- Member, Editorial Board, Web Semantics.

2002- Member, Editorial Board, Web Intelligence and Agent Systems (WIAS).

1999- Member, Editorial Board, Autonomous Agents and Multi-Agent Systems

e. Collaborators & Other Affiliations

Acquisti, Alessandro. Carnegie Mellon University, Pittsburgh, PA; Allen, James, IHMC/UWF and Univ. of Rochester; Ambrose, Rob, NASA-JSC; Andrasik, Frank, IHMC/UWF; Beautement, Patrick, QinetiQ, Ltd., Malvern, England; Bradshaw, Jeffrey M., IHMC/UWF; Boy, Guy, EURISCO, France; Breedy, Maggie, IHMC/UWF; Brinn, Marshall, BBN, Cambridge, MA; Bunch, Larry, IHMC/UWF; Burstein, Mark, BBN, Cambridge, MA; Canas, Alberto, IHMC/UWF; Carvalho, Marco M., IHMC/UWF; Chambers, Nate, IHMC/UWF; Clancey, Willliam, NASA-Ames Research Center; Cranfill, Rob, Boeing Corp., Seattle, WA; Drakunov, Sergey V., Tulane University; Duke, Mike, Colorado School of Mines; Durfee, Ed, University of Michigan; Etzioni, Oren, University of Washington; Feltovich, Paul and Joan, IHMC/UWF; Fikes, Richard, Stanford University; Ford, Kenneth, IHMC/UWF; Galescu, Lucian, IHMC/UWF; Gawdiak, Yuri, NASA-JSC; Gray, Robert, Dartmouth University; Greaves, Mark, Defense Advanced Research Projects Agency; Gruninger, Michael, NIST; Guedry, Frederick E., IHMC/UWF; Holmback, Heather, Boeing Corp., Seattle, WA; Hewitt, Rattikorn, IHMC/UWF; Hexmoor, Henry, University of Arkansas; Hoberman, Chuck Hoberman Associates, Inc., New York; Hoffman, Robert R., IHMC/UWF; Jeffers, Renia, IHMC/UWF; Jensen, Wayne, NIST; Johnson, Matthew, IHMC/UWF; Jung, Hyuckchul, IHMC/UWF; Kass, Steven J., IHMC/UWF; Kerstetter, Mike, Boeing Corp., Seattle, WA; Kulkarni, Shriniwas, IHMC/UWF; Lawrence, Craig, IDEO Corp.; Lieberman, Henry, MIT, Cambridge, MA; Lott, James., IHMC/UWF; Rachid Manseur, University of West Florida; McGuiness, Deborah, Stanford University; Montanari, Rebecca, University of Bologna, Italy; Neves, Ken, Boeing Corp., Seattle, WA; Perry, James F., IHMC/UWF; Pratt, Jerry, Yobotics, Inc., Boston, MA; Rupert, Angus H., Naval Aerospace Medical Research Laboratory; Schlegel, Todd T., NASA- Johnson Space Center; Sierhuis, Maarten, RIAC; Silverman, Barry, University of Pennsylvania; Sullivan, Keith, Duke University; Suri, Niranjan, IHMC/UWF; Tate, Austin, University of Edinburgh; Taysom, William., IHMC/UWF; Tonti, Gianluca, University of Bologna, Italy; Uschold, Mike, Boeing Corp., Seattle, WA; Uszok, Andrej, IHMC/UWF

Graduate and Postdoctoral Advisors: Earl "Buz" Hunt, University of Washington
Thesis Advisor and Postgraduate-Scholar Sponsor: Luc Haudot, Ecole Nationale Superieure de l'Aviation et de l'Espace, Toulouse, France; Jean Koning, Universite de Paul Sabatier, Toulouse, France; Gerald Knoll, University of West Florida; Monique Calisti, ETH, Lausanne, Switzerland; Thomas Cowin, University of West Florida; Marco Carvalho, University of West Florida; Shawn R. Murray, University of West Florida

Patrick J. Hayes, Ph.D. **phayes@ihmc.us**

Professional Preparation

University of Cambridge, B.A. (Math tripos), 1966
University of Edinburgh, Ph D. (Artificial Intelligence) 1973

Appointments

Research Scientist, Institute for Human and Machine Cognition (IHMC), and John C. Pace Eminent Scholar, University of West Florida (Sept 1996 -)
Research Professor, Depts. of Computer Science and Philosophy, and Beckmann Institute, University of Illinois at Urbana-Champaign (Sept 1992 – Sept 1996).
Senior Member of Technical Staff, Microelectronics and Computing Technology Corporation, Austin, Texas; Director of CYC-West project, Palo Alto, Ca. (Feb 1991 – Sept 1992).
Principal Scientist, Xerox-PARC (1987-1990)
Senior Member of Technical Staff, Schlumberger Palo Alto Research Center (1985 – 87)
Visiting Scholar, CSLI, Stanford (January 1985; Sept 1987 –)
Consulting Professor, Dept. of Computer Science, Stanford (1985-1994)
Luce Professor of Cognitive Science, Depts of Computer Science, Philosophy and Psychology; Chair, Cognitive Sciences Cluster, University of Rochester (1981- 1985)
Fellow, Center for Advanced Study in the Behavioral Sciences (CASBS), Stanford (1979-1980)
Professeur Invité, Université de Geneve (Jan 1978 – April 1978)
Lecturer, then Senior Lecturer, Dept. of Computer Science, University of Essex, UK. (1973-79; 1979-80)

Related Publications

Carroll, J., Bizer, C., Hayes, P., Stickler, P. (2005) *Named Graphs, Provenance and Trust* (HP technical report HPL-2004-57 (2004) http://www.hpl.hp.com/techreports/2004/HPL-2004-57.pdf; accepted for 14th International WWW conference, Keio, Japan 2005)

Hayes, P. *RDF Semantics* (W3C Recommendation 10 February 2004) http://www.w3.org/TR/2004/REC-rdf-mt-20040210/

Patel-Schneider, P., Hayes, P., Horrocks, I. *OWL Web Ontology Language: Semantics and Abstract Syntax* (W3C Recommendation 10 February 2004) http://www.w3.org/TR/2004/REC-owl-semantics-20040210/

Menzel, C., Hayes, P. SCL: A Logic Standard for Semantic Integration. (2003) *Proc. Workshop on Semantic Integration, Second International Semantic Web Conference* (ISWC 2003)

Bradshaw, J. M., Uszok, A., Jeffers, R., Suri, N., Hayes, P., Burstein, M. H., Acquisti, A., Benyo, B., Breedy, M. R., Carvalho, M., Diller, D., Johnson, M., Kulkarni, S., Lott, J., Sierhuis, M., & Van Hoof, R. (2003). Representation and reasoning for DAML-based policy and domain services in KAoS and Nomads. *Proceedings of the Autonomous Agents and Multi-Agent Systems Conference (AAMAS 2003)*. 14-18 July, Melbourne, Australia. New York, NY: ACM Press, pp. 835-842.

Significant Publications

Hayes, P. (1995) "A Catalog of Temporal Theories", *Tech report UIUC-BI-AI-96-01*, University of Illinois

Hayes, P. (1985) "Naive Physics I: Liquids" in J.R. Hobbs and R.C. Moore (Eds.) *Formal Theories of the Common Sense World* (Vol. 1). Norwood, NJ: Ablex Publishing Company, 1985.

[Reprinted in: *Readings in Cognitive Science*, ed. Collins and Smith, 1988.
Readings in Qualitative Reasoning about Physical Systems, ed. Weld and deKleer, 1989.]

Hayes, P. (1980) "The Logic of Frames" in D. Metzing (Ed.) *The Frame Reader*. DeGruyter, Berlin, 1980.

[Reprinted in: *Readings in Artificial Intelligence,* ed. Webber & Nilsson, Tioga 1982
Readings in Knowledge Representation, ed. Brachman and Levesque, MorganKaufmann 1985.]

Hayes, P. (1974) "Some Problems and Non-Problems in Representation Theory"' *Proc., 1st AISB Conference,* U. Sussex, 1974.

[Reprinted in: *Readings in Knowledge Representation*, ed. Brachman and Levesque MorganKaufmann 1985.]

McCarthy, J. Hayes, P. (1969) "Some philosophical problems from the standpoint of Artificial Intelligence", *Machine Intelligence 4*, Edinburgh University Press. 1969.

[Reprinted in: *Readings in Artificial Intelligence*, ed. Webber & Nilsson, Tioga 1982
Readings in Planning, ed. Allen, Hendler &Tate, MorganKaufmann 1990.
Formalizing Common Sense, ed. Vladimir Lifschitz, Ablex 1990]

Synergistic Activities

1. Wrote what is arguably the first major effort in applied ontology (Hayes 1985), cited as the justification for election as Charter Fellow of Cognitive Science Society, 2001.
2. At various times have been involved in many academic/professional organizations, including secretary of AISB (1968-79; UK national society), Chair of IJCAI (1980-84; world body organizing biennial international conference sequence), President of AAAI (1991-93; US national professional organization. Appointed new executive director, managing editor of AAAI Press, began Fellows program and initiated major public-education initiatives), member of governing board of Cognitive Science Society (1983-86). Associate editor of the major journal in the field (AI Journal 1979-86) and on the editorial boards of J. Logic Programming, J. Cognitive Science, J. of Applied AI, and founding editor of two book series ("Symbolic Computation", Springer-Verlag 1981; "Computational Models of Thought and Language", MIT Press 1983).
3. While Luce Professor at Rochester, oversaw the development of the first undergraduate degree program in Cognitive Science to be accredited at a major US University, coordinating efforts across seven departments in four Schools and creating five new courses. The program was accredited by the State of NY in 1985 and was continued through 2002; it has since been absorbed into a new academic department of Brain and Cognitive Sciences.
4. Since 2001, actively involved with a number of standard-setting efforts connected with the semantic web initiative as a W3C Consortium invited expert. I have been a member of four W3C Working Groups and co-authored two Recommendations (RDF and OWL, the chief semantic web description languages) and am a member of the DAML Joint Committee which designed DAML+OIL, the precursor to OWL and SWRL, the semantic web rule language; and of DAWG, which is designing SPARQL, the first query language for the semantic web, and which is partly based on DQL, designed by myself, Richard Fikes and Ian Horrocks.
5. Elected Fellow of American Association for Artificial Intelligence, Cognitive Science Society.

Collaborators & Other Affiliations (last 5 years)

Murray Altheim, Open University, UK
Christian Bizer, Freie Universitat Berlin
Jeff Bradshaw, Ph.D., IHMC
Alberto Cañas, Ph.D., IHMC and UWF
Jeremy Carroll, Ph.D., Hewlett Packard Labs, Bristol UK
Dan Connolly, Ph.D, World Wide Web Consortium and MIT
Vinay Chaudhri, Ph.D., SRI International
Prof. Max Egenhofer, U. Maine
Prof. Richard Fikes, Stanford
Paul Feltovitch, Ph.D., IHMC
Kenneth Ford, Ph.D., IHMC and UWF
Ramanathan. V. Guha,, PhD., IBM Corp
Sandro Hawke, World Wide Web Consortium and MIT
Robert R. Hoffman, Ph.D,, IHMC
Prof. Kathleen Hornsby, U. Maine
Prof. Ian Horrocks, Manchester University, UK
Prof. Robert Kowalski, Imperial College, UK (rtd.)
Mala Mehrotra, Ph.D., Pragati, Inc.
Prof. Deborah McGuinness, KSL, Stanford
Prof. Christopher Menzel, Texas A&M U.
Anil Raj, IHMC
Peter Patel-Schneider, Ph.D., Lucent Laboratory
Paolo Pinheira da Silva, KSL, Stanford
Patrick Stickler, Nokia Research Center, Finland

Graduate advisors

Prof. Donald Michie, Prof. Bernard Meltzer, both University of Edinburgh.

Paul J. Feltovich, Ph.D.

Professional Preparation

Allegheny College, Meadville, PA	Mathematics	B.S., 1969
University of Minnesota	Counseling Psychology	Doctoral Program, 1969-1973
University of Minnesota	Educational Psychology	Ph.D., 1981
University of Pittsburgh	Cognitive Psychology	Post-Doctoral Fellow, 1978-1982

Appointments

Research Scientist, Institute for Human and Machine Cognition, University of West Florida (effective September, 2001)

Interim Chair, Dept. of Medical Education, Southern Illinois University School of Medicine (Sept.1, 1999-Sept.1, 2001).

Adjunct Professor, Dept. of Educational Psychology, University of Illinois at Urbana-Champaign, (1998-present).

Professor, Dept. of Medical Education, SIU School of Medicine (1997-2001).

Director, Cognitive Science Division, Dept. of Medical Education, SIU School of Medicine (1988-2001).

Associate Professor, Dept. of Medical Education, SIU School of Medicine (1988-1997).

Assistant Professor, Dept. of Psychiatry, SIU School of Medicine (1986-2001).

Assistant Professor, Dept. of Medical Humanities, SIU School of Medicine (1986-1990).

Assistant Professor, Dept. of Medical Education, SIU School of Medicine (1982-1988).

Related Publications

Klein, G., Feltovich, P.J., Bradshaw, J.M., & Woods, D. D. (in press, 2005). Common ground and coordination in joint activity. In W.R. Rouse & K.B. Boff (Eds.), *Organizational simulation*. New York: Wiley.

Feltovich, P.F., Hoffman, R.R., Woods, D.W., & Roesler, A. (2004). Keeping it too simple: How the reductive tendency affects cognitive engineering. *IEEE Intelligent Systems, 19(3),* 90-94.

Feltovich, P.J., Bradshaw, J.M., Jeffers, R., Suri, N. & Uszok, A. (2004). Social order and adaptability in animal and human cultures as analogues for agent communities: Toward a policy-based approach. In Omacini, A., Petta, P., & Pitt, J. (Eds.), *Engineering societies for agents world IV* (pp.21-48). Lecture Notes in Computer Science Series. Heidelberg, Germany: Springer-Verlag.

Bradshaw, J. M., Feltovich, P. J., Jung, H., Kulkarni, S., Allen, J., Bunch, L. Chambers, N. Galescu, L. Jeffers, R., Johnson, M. Sierhuis, M., Taysom, W., Uszok, A. & Van Hoof, R. (2004). Policy-based coordination in joint human-agent activity. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics,* The Hague, The Netherlands, 10-13 October..

Feltovich, P.J., Spiro, R.J. & Coulson, R.L. (1997). Issues of expert flexibility in contexts characterized by complexity and change. In P. Feltovich, K. Ford, & R. Hoffman (Eds.), *Expertise in context: Human and machine* (pp. 125-146), Cambridge, MA: AAAI (American Association for Artificial Intelligence)/MIT Press.

Significant Publications

Feltovich, P., Spiro, R., & Coulson, R. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In D. Evans & V. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 113-172). Cambridge, MA: MIT Press.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., & Klopfer, D. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 311-342). Hillsdale, NJ: Lawrence Erlbaum Associates.

Spiro, R., Coulson, R., Feltovich, P., & Anderson, D. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. In *Proceedings of the 10th Annual Conference of the Cognitive Science Society* (pp. 375-383). Hillsdale, NJ: Lawrence Erlbaum Associates.

Feltovich, P., Johnson, P., Moller, J., & Swanson, D. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W. Clancey & E. Shortliffe (Eds.), *Readings in medical artificial intelligence: The first decade* (pp. 275-319). Reading, MA: Addison Wesley.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152.

Synergistic Activities

1. Major contributor to the modern understanding of human expertise, with an article in this field that is the most highly cited article in the history of the *Cognitive Science* journal and a designated Science Citation Classic (Institute for Scientific Information).
2. Inventor and co-inventor of two cognitive constructs, the Logical Competitor Set and the Illness Script, that have had major influence in several fields, including psychology, medical education and diagnosis, and artificial intelligence.
3. Co-inventor of the construct of the Reductive Bias, the inclination for learners and practitioners to oversimplify complex material in systematic ways, that has had influence on learning, instruction and, progressively, in real world applications such as workplace design, safety and error.
4. Co-developer of Cognitive Flexibility Theory, a major modern theory of learning in complex and ill-structured knowledge domains.
5. Major developer and trainer for applications of Problem-Based Learning (a small group and case-based method of learning) in fields inside of, but also outside of, medicine where PBL was first developed.

Collaborators

Howard Barrows, MD., Southern Illinois University School of Medicine
Jeffrey Bradshaw, Inst. Human and Machine Cognition
Richard Coulson, Ph.D., Southern Illinois University
Neil Charness, Ph.D., Florida State University
Sharon Derry, Ph.D., University of Wisconsin
David Eccles, Florida State University
Anders Ericsson, Ph.D., Florida State University
Steve Fiore, University of Central Florida
Ken Forbus, Ph.D., Northwestern University
Ken Ford, Ph.D., Inst. Human and Machine Cognition
Jack Hansen, Ph.D., Inst. Human and Machine Cognition
Robert R. Hoffman, Ph.D. Inst. Human and Machine Cognition
Renia Jeffers, Ph.D., Inst. Human and Machine Cognition
Gary Klein, Ph.D., Klein Assoc., Dayton, OH
Michael J. Prietula, Ph.D., Florida International University
Rand Spiro, Ph.D., Michigan State University
Niranjan Suri, Ph.D., Inst. Human and Machine Cognition
Andrzej Uszok, Ph.D., Inst. Human and Machine Cognition
David Woods, Ph.D., Ohio State University

Doctoral and Post Doctoral Advisors

Micki Chi, Ph.D., University of Pittsburgh (Post Doctoral)
Robert Glaser, University of Pittsburgh (Post Doctoral)
Paul E. Johnson, Ph.D., University of Minnesota (Doctoral)
Alan Lesgold, Ph.D., University of Pittsburgh (Post Doctoral)

Thesis Advisor & Postgraduate-Scholar Sponsor

Sanjeev Dutta, MEd., M.D., Stanford University School of Medicine
David Eccles, Ph.D., Florida State University