Title of Proposal and name of PI:

Using Crowdsourcing To Measure Complex Social Concepts: A Pilot Study
Principal investigator: Glenn Palmer

Descriptive Keywords:

measurement, social science, crowdsourcing, military conflict, document classification

Abstract:

Much of the data we collect in the social sciences come from news reports made available by providers such as *LexisNexis*. The standard approach for collecting data from these sources is to query their database, retrieve a set of documents, and use trained, expert coders to manually read through the returns in search of relevant information. The manual search and the necessity of expert coders are the primary sources of inefficiency in this process. In the proposed pilot study, our inter-disciplinary team will apply automated document classification and crowdsourcing to address these inefficiencies when coding instances of international military conflict: a complex social concept. Although our experiments will involve international military conflict, the method has broad applicability across disciplines in the social sciences. We begin by improving the precision of the retrieved document set using machine learning techniques for automated document classification. Then, the remaining documents are supplied to a crowd of workers through a platform such as Amazon's Mechanical Turk. Each document is read by a worker who is paid a small amount of money to answer a dynamic line of simple, objective questions designed to remove the necessity of expertise from the process. We request funds for the purposes of developing the techology for classifying the documents and to experiment with a crowdsourcing platform that is capable of handling the dynamic line of questioning asked of each coder. The Militarized Interstate Dispute project, a dataset on international military conflict which is housed at Penn State and has been awarded three NSF grants in the past six years, has provided the document sets to be classified and crowdsourced as well as the results of the content analysis by expert coders. The results obtained from crowdsourcing are compared to those of the expert coders to assess the accuracy of the method.

Specific Aims and Objectives

While distributed labor (i.e. crowdsourcing) has been used for the purposes of measurement in social science, the conventional wisdom is that its potential for measuring complex social concepts is limited and that such concepts may only be measured by trained experts. Our primary objective is to demonstrate that crowdsourcing *can* be used to measure complex social concepts *without* the use of trained experts in the coding phase.[1] This SSRI Level II application will launch a pilot study to begin assessing that objective in two ways. The first is to further develop the automated document classification system initially developed by the Militarized Interstate Dispute project in an effort to improve the precision – the ratio of true positives to all documents classified as positive – of the document set to be crowdsourced. The second is to support a series of crowdsourcing experiments using platforms such as Mechanical Turk.[2] Although our experimental application is to international military conflict, the method is applicable to any situation where trained experts pore over a set of documents to extract relevant information in a process known generally as content analysis (Krippendorff 2013). All the technology developed over the course of this project is to be open source and we intend to publish the results of our experiments in a general social science journal. Should the experiments prove successful, we intend to submit an NSF proposal for the purposes of further developing the method and applying it to the next iteration of the Militarized Interstate Dispute project.

Brief background, rationale, and description of the methods:

Measuring complex social concepts is a fundamental task in all of quantitative social science (Krippendorff 2013; Goertz 2006; Sartori 1984; Zeller and Carmines 1980). Much of the information used to measure these concepts comes from the content analysis of news reports. This has traditionally entailed using trained expert coders to read over the reports and code relevant information into data. In this pilot study our goal is to conduct experiments designed to determine the ability of crowdsourcing to replace these trained experts.

  Crowdsourcing is a term that refers to the utilization of distributed labor from an

---

[1]In our context, distributed labor, or crowdsourcing, refers to a process where workers are paid small amounts of money to carry out simple tasks that require human intelligence (as opposed to artificial intelligence), such as image detection (Nowak and Rüger 2010) or, in our case, content analysis.

[2]Mechanical Turk is a crowdsourcing platform developed by Amazon for the purposes of linking requesters (us) with workers (the crowd). More information about crowdsourcing and Amazon's Mechanical Turk can be found at `https://www.mturk.com/mturk/`

online community for the purposes of carrying out simple and often redundant tasks. The online community is accessed through one of the many crowdsourcing platforms, such as Amazon's Mechanical Turk or Crowdflower. Through such platforms, a task is posted to the website and workers, who voluntarily log in and search through posted tasks, are paid a small amount of money to complete the task.

A variety of experiments have demonstrated the utility of crowdsourcing in solving problems that were previously thought to require substantive expertise. For example, many natural language processing tasks, such as various types of text annotation, are amenable to the use of crowdsourcing in place of trained experts (Zhai et al. 2012; Sabou, Bontcheva and Scharl 2012; Snow et al. 2008). With proper incentives in a web-based gaming platform, non-experts have been shown to be capable of predicting protein structures (Cooper et al. 2010).

In the social sciences, crowdsourcing has seen its most widespread use in the area of collecting sensor data on human populations (Aggarwal and Abdelzaher 2011).[3] For example, in times of crisis, such as a natural disaster, crowdsourcing can be used to aid relief efforts (Goodchild and Glennon 2010; Zook et al. 2010). In general, crowdsourcing can be used for collecting many types of geospatial data that was previously the exclusive domain of international and transnational organizations (Heipke 2010).

While some researchers have begun working with crowdsourcing to measure social concepts (Benoit et al. 2012; Honaker 2012), the complexity of the concepts capable of being crowdsourced is largely unknown. In a very basic sense, the conventional wisdom is that the more complex the concept, the greater the expertise needed for coding and the less likely crowdsourcing is to be of use. Our experiments challenge this wisdom by deconstructing a complex social concept down to its simple, fact-based component parts and asking workers to answer objective questions detailing those parts.

Our experimental concept will be international military conflict as defined by the Militarized Interstate Dispute (MID) project, a data collection project housed at Penn State. The MID dataset is one of the most commonly used datasets in quantitative political science, and it is complex.[4] Each MID is comprised of individual events, called militarized international incidents (MIIs). Each MII is a specific threat, display, or use of military force by one country against another (Ghosn, Palmer and Bremer 2004).

Our research design proceeds in three not necessarily sequential phases, beginning with a set of documents provided by the MID project. Importantly, these documents

---

[3]Sensor data in this sense is essentially the location of an individual person at a given point in time.

[4]At the time of this writing, the first three iterations of the MID project have generated 519, 796, and 462 citations on Google Scholar, respectively.

have already been coded by expert coders, thus providing us with an ideal test case for crowdsourcing experiments. The first phase consists of utilizing automated document classification techniques for subsetting out documents that are not relevant to the MID. The fourth iteration of the MID project has incorporated automated document classification into its data collection method and this system will provide our starting point (D'Orazio et al. 2012). For classifying the documents, the MID project has used both inductive and transductive support vector machines, implemented with Thorsten Joachims' SVM^{light} software (Joachims 2002).

Phase Two consists of deconstructing the MID down to its most basic, objective parts. The idea is to move from something complex to something simple so that *expertise is not necessary for coding*. As a result, various crowdsourcing techniques become possible because of the simplicity of the component parts. Phase Three consists of implementing crowdsourcing to code the factual data on these basic components of the concept.

For the crowdsourcing phase, workers begin by reading a news story, generally between one and six paragraphs in length. Then, workers will be prompted to answer some very general questions whose answers are then used to determine the increasingly targeted questions that will be asked next. If we think of deconstructing a complex concept into a decision tree, users will proceed down the tree depending on their responses. When the last node in the tree has been reached, all meaningful information will have been extracted from the document.

It is important to note that the questions asked to the workers are objective, fact-based questions. The questions might include such things as "Is the military of South Korea directly involved in this paragraph?" or "We understand this document is about a mlitary conflict between Turkey and Syria, what is the precise location of this incident?" Such questions require a certain level of reading competency on the part of the workers but do not require any sort of expertise in the subject matter. Mechanical Turk and Crowdflower are the two platforms we will begin experimenting with because of their capability of handling the dynamic line of questioning that will be asked of each coder and because platforms such as these have mechanisms for identifying high-competency workers.

The documents that will be supplied to the crowd have already been coded by trained expert coders working on the MID project. The time it took the experts to code these documents, as well as the cost of the MID project as a whole, is already known, thus making it an ideal test case. These document sets are used in our experiments, and the data retrieved through crowdsourcing are directly compared to the results from the

trained experts.

Relevance to the SSRI:

The SSRI encourages inter-disciplinary research with a clear contribution to the social sciences. Our team, which consists of faculty from Communications, Information Sciences and Technology, and Political Science are collaborating to develop a method of measurement with widespread applicability in the social sciences. We expect our pilot study to lead to a larger project funded by the NSF.

Anticipated Outcomes:

Upon completion of this pilot study we will have conducted one of the earliest experiments utilizing crowdsourcing for content analysis. International military conflict is a complex social concept and therefore a difficult test case. By demonstrating its ability collect data here, we believe we will be providing a firm basis for future projects using crowdsourcing for other concepts. We will present the results of our experiments at academic conferences, including the meetings of the Peace Science Society, the Society for Political Methodology, and the Association for the Advancement of Artificial Intelligence. Furthermore, we intend to publish our experiments in a general social science journal. We expect our results to support the notion that crowdsourcing can be used to measure complex social concepts. Should this hold, we anticipate pursuing an NSF grant for the purposes of developing the methodology and applying it to the expansion of the MID data from 2011-2016.

Part II: Additional Information

External Funding Sources:

We are reasonably optimistic about our chances of securing funding from NSF, provided the project's method should prove sound. We have been in contact with Brian Humes, the Program Manager for political science at NSF, and his assistant, Erik Herron. Mr. Herron has responded that "as long as you link the project to critical theoretical questions in the literature, then it could be a match [for political science]." He also notes that "it might also be (more) appropriate for the MMS program or possibly the interdisciplinary solicitation IBSS," an observation we will take into consideration should the effects of the Coburn Amendment be felt beyond September. Mr. Herron had a couple of constructive suggestions to make. For one, he said "you may want to refine how you address the

theoretical literature to clearly define the advances that the project will inspire and how it may be generalizable beyond the case under investigation."

It is worth saying Herron's responses are very general and do not indicate any familiarity with past NSF-supported research the MID project has earned; Brian Humes is much more conversant about the project and its past NSF grants. Those include: eleven coordinated grants for the MID3 Project, completed in 2003; a grant to investigate the efficiency of the last method employed for the MID update, 2002-2010, for $143,631; the MID4 data-gathering project itself, for $316,140; and an informally approved grant to complete the project for $71,781. (This last grant is currently on hold as NSF seeks to learn how to deal with the Coburn Amendment.)[5]

Timeline:

Due to our use of human participants through crowdsourcing, we have submitted our project to Penn State's Institutional Review Board. IRB has determined our project is exempt from review, both initial and ongoing.[6]

The crowdsourcing experiments for this project are to be completed in June and July, 2013. This includes the initial experiments for developing and improving both the automated and crowdsourced components, as well as the primary experiment which collects a year's worth of data for the purposes of comparing the crowdsourced data to the expert-coded data.

With the experiments conducted, we will make all technology developed available for others to use freely. We expect to submit the results of these experiments for publication by October and to present the results at various conferences over the course of the 2013-2014 academic year.

The NSF proposal will be submitted in winter of 2014. Pending approval, we expect work on this project to begin in fall of 2014.

Personnel: Glenn Palmer, Vito D'Orazio, David Reitter, Michael Schmierbach

Glenn Palmer is a professor of political science at Penn State and is the PI for this project. In addition to overseeing the project as a whole, he has a specific role with respect to the MID data. Palmer has been PI or Co-PI on the two most recent MID projects and

---

[5]NSF award numbers 0002568, 0719634, and 0924240, respectively.
[6]The IRB approval letter is appended to this application.

will validate the accuracy of the dynamic questions being asked of coders as well as the answers provided by the coders.

Vito D'Orazio is a PhD candidate in political science at Penn State and will be a collaborating investigator on this project. For the 2013-2014 academic year, he will be on a fellowship at the Institute for Quantitative Social Science at Harvard University. He will be coordinating work between all project personnel as well as working specifically on the software necessary for implementing the machine learning algorithms for classification and named entity recognition.

David Reitter is an assistant professor of Information Sciences and Technology at Penn State. As a collaborating investigator on this project, he will be working primarily on the software supporting the crowdsourcing. He has a background in software development and computational linguistics (as the author of Aquamacs Emacs). Dr. Reitter's research uses cognitive modeling to describe cognitive processes at the interface between individual and groups. His substantive interest in this project pertains to the synergies between individual worker's cognitive task and the network structure allowing work pieces to be combined.

Michael Schmierbach is an assistant professor of communcations at Penn State and an expert in content analysis. Schmierbach will be a collaborating investigator on this project. He will be working primarily on developing the generalized content analysis framework in addition to assessing variation in media content across sources and the potential bias associated with such variation.

Other personnel include three additional research assistants, two of which are experienced with computer programming and at least one of which is experienced with the MID data. The RA experienced with the MID data will work primarily with Palmer to construct the dynamic line of questions that will be asked of the workers. The other RAs will work with D'Orazio and Reitter to write the software for the crowdsourcing implementation and classification improvements.

Budget and Justification:

Budget Summary
- Initial experiments: worker compensation
  - 3,000 dollars
- Primary experiment: worker compensation
  - 7,873 dollars
- Research assistants
  - 9,127 dollars

- Total request
  - 20,000 dollars

The majority of the funding is being requested to pay the workers who will be answering the questions in our experiments. Based on what has been paid to them by (Benoit et al. 2012), and based on their expected workload and what is considered to be fair pay, we expect to pay our workers sixty cents per document. The initial crowdsourced experiments will be conducted with 500 to 1,000 documents. We expect five rounds of initial experiments to be sufficient for fine-tuning the process prior to conducting the primary experiment. Although the exact number depends on the initial results, we expect the initial experiments to cost about 3,000 dollars.

The primary experiment is the the evaluation of documents about military conflicts from the year 2010 that will be provided by the MID project. The MID project has paid graduate research assistants to read through 12,946 documents from the year 2010 and stucture data on MIDs based on these reports. We expect to run a complete evaluation of these documents using the new approach and compare the crowdsourced results to the expert results. The expected cost is 7,873 dollars and includes a minor additional cost of 105 dollars for providing refreshments at meetings for project personnel.

We request the remaining funds for RAs (including D'Orazio) to implement and assist in analyzing the information extraction, document classification, and crowdsourcing experiments. D'Orazio is to work thirty hours per week for the first six weeks of the project (or until July 1, at which point he will be on fellowship) at a cost of twelve dollars per hour (2,160 dollars). David Reitter's RA is to work for at least the first half of the summer at a cost of 2,820 dollars. The remaining funds are for RAs to work primarily with Palmer constructing the dynamic questions and analyzing the results. All estimates for RA support include fringe benefits at a rate of 7.8%, as specified in the budget proposal form.

SSRI Services:

We do not expect additional SSRI services will be required for this project.

Investigator Information:

Lead Investigator:

Glenn Palmer

Professor

Department of Political Science

College of Liberal Arts, University Park

814-865-5594

gpalmer@psu.edu

Collaborating Investigator:

Vito D'Orazio

PhD Candidate

Department of Political Science

College of Liberal Arts, University Park

516-640-8937

vjd125@psu.edu

Collaborating Investigator:

David Reitter

Assistant professor

Information Sciences and Technology

College of Information Sciences and Technology, University Park

412-417-0529

reitter@psu.edu

Collaborating Investigator:

Michael Schmierbach

Assistant professor

Department of Media Studies

College of Communications, University Park

814-865-9582

mgs15@psu.edu

Pre-Submission Checklist Questions

Q.1 Which agency or foundation officials (e.g., project officer) have you spoken with to determine their interest in this project or project area? What feedback did you receive on your concept and approach?

A.1 We approached Brian Humes, Program Officer in political science at the National Science Foundation, and his assistant, Erik Herron. Herron's response, in part, was:

> "You asked if the project would be appropriate for Political Science and the likelihood of funding. In terms of the former question, as long as you link the project to critical theoretical questions in the literature, then it could be a match. Given the description, it might also be (more) appropriate for the MMS program or possibly the interdisciplinary solicitation IBSS."

Herron said also that,

> "To enhance competitiveness, you may want to refine how you address the theoretical literature to clearly define the advances that the project will inspire and how it may be generalizable beyond the case under investigation."

Q.2 Are you responding to a specific request for proposal (RFP/RFA), program announcement, or other special funding initiative? If yes, which one and how is your Level 2 a good match for it?

A.2 No

Q.3 3. Is this Level 2 being undertaken in response to feedback from a prior external proposal? If so, how does this project address reviewer concerns?

A.3 No

Q.4 How does your study compare with projects in similar domains that have been funded by your targeted agency? In particular, how does the scope of your methodology appear similar to other funded projects (in terms of the size and representativeness of the sample, measurement strategies, design and planned analytic approach, etc.)?

A.4 Our pilot study builds on the technology and theoretical relevance of previous NSF-funded Militarized Interstate Dispute (MID) projects (NSF award numbers 0924240, 0719634, 0002568). We test a research design that utilizes distributed labor (i.e. crowdsourcing) for the purposes of improving the data collection process. Two examples of related studies that seek to apply crowdsourcing for similar purposes and have been funded by the NSF are at Harvard University (and collaborating universities) for the purposes of analyzing historical information (NSF award number 1244667) and at Ohio State University and the University of Washington for the purposes of collecting geographic information (NSF award numbers 1048100 and 0849625). In comparison to these NSF-funded projects, our study utilizes a similar crowdsourcing framework but seeks to apply it to the measurement of complex social concepts.

Q.5 What criteria will be used to evaluate your proposal and what do you know about the likely reviewers?

A.5 Erik Herron spoke directly to this point when we contacted him: "Programs at the NSF, including Political Science, evaluate all proposals according to our primary merit criteria: Intellectual Merit and Broader Impacts. You can read more about the merit criteria and process in the NSF Grant Proposal Guide (updated in January 2013 http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf13001). If you were to submit your project to Political Science, the ad hoc reviewers and panel would assess it based on the main merit criteria."

Q.6 What input/advice/support have you received from your department head and/or college research dean?

A.6 Our head, Scott Bennett, is aware of the proposed project and is very supportive. He believes, as we do, that data will have to be coded much more efficiently in the future and that what is being proposed here is a promising avenue to investigate. Professor Bennett is also aware that we are submitting this proposal.

References

Aggarwal, Charu C. and Tarek Abdelzaher. 2011. Integrating Sensors and Social Networks. In *Social Network Data Analytics*, ed. Charu C. Aggarwal. Springer US pp. 379–412.
URL: *http://dx.doi.org/10.1007/978-1-4419-8462-3_14*

Benoit, Kenneth, Drew Conway, Michael Laver and Slava Mikhaylov. 2012. "Crowdsourced data coding for the social sciences: massive non-expert human coding of political texts.". Paper presented at the Meeting of the Midwest Political Science Association.

Cooper, Seth, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popovic and Foldit Players. 2010. "Predicting Protein Structures With a Multiplayer Online Game." *Nature* 466(August):756–760.

D'Orazio, Vito, Steven T. Landis, Glenn Palmer and Philip Schrodt. 2012. "Separating the Wheat from the Chaff: Applications of Automated Document Classification to MID.". Presented at the MidWest Political Science Meeting, 2011. Available at `http://vitodorazio.weebly.com/papers.html`.

Ghosn, Faten, Glenn Palmer and Stuart A. Bremer. 2004. "The MID3 Data Set, 1993-2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21(2):133–154.

Goertz, Gary. 2006. *Social Science Concepts: A User's Guide*. Princeton, NJ: Princeton University Press.

Goodchild, Michael F. and J. Alan Glennon. 2010. "Crowdsourcing Geographic Information for Disaster Response: A Research Frontier." *International Journal of Digital Earth* 3(3):231–241.

Heipke, Christian. 2010. "Crowdsourcing Geospatial Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 65(6):550–557.

Honaker, James. 2012. "Mechanical Turk for Text Annotation and Judgment Tasks.". Talk sponsored by Penn State's Quantitative Social Science Initiative.

Joachims, Thorsten. 2002. *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms.* Norwell, MA: Kluwer Academic Publishers.

Krippendorff, Klaus. 2013. *Content Analysis: An Introduction to its Methodology.* Third ed. Thousand Oaks, CA: Sage.

Nowak, Stefanie and Stefan Rüger. 2010. How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the international conference on Multimedia information retrieval.* MIR '10 New York, NY, USA: ACM pp. 557–566.
URL: *http://doi.acm.org/10.1145/1743384.1743478*

Reitter, David. 2003. "Simple signals for complex rhetorics: On rhetorical analysis with rich-feature support vector models." *Journal for Language Technology and Computational Linguistics* 18(1/2):38–52.

Sabou, Marta, Kalina Bontcheva and Arno Scharl. 2012. Crowdsourcing research opportunities: lessons from natural language processing. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies.* i-KNOW '12 New York, NY, USA: ACM pp. 17:1–17:8.
URL: *http://doi.acm.org/10.1145/2362456.2362479*

Sartori, Giovanni. 1984. *Social Science Concepts: A Systematic Analysis.* Beverly Hills: Sage.

Snow, Rion, Brendan O'Connor, Daniel Jurafsky and Andrew Y. Ng. 2008. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing.* EMNLP '08 Stroudsburg, PA, USA: Association for Computational Linguistics pp. 254–263.
URL: *http://dl.acm.org/citation.cfm?id=1613715.1613751*

Zeller, Richard A. and Edward G. Carmines. 1980. *Measurement in the Social Sciences: The Link Between Theory and Data.* New York: Cambridge University Press.

Zhai, Haijun, T. Lingren, L. Deleger, Qi Li, M. Kaiser, L. Stoutenborough and I. Solti. 2012. Cheap, Fast, and Good Enough for the Non-biomedical Domain but is It Usable for Clinical Natural Language Processing? Evaluating Crowdsourcing for Clinical Trial Announcement Named Entity Annotations. In *Healthcare Informatics, Imaging and*

Systems Biology (HISB), 2012 IEEE Second International Conference on. pp. 106–106.

Zook, Matthew, Mark Graham, Taylor Shelton and Sean Gorman. 2010. "Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake." *World Medical & Health Policy* 2(2):7–33.
URL: *http://dx.doi.org/10.2202/1948-4682.1069*

Vito D'Orazio
221 Pond Lab
Pennsylvania State University
University Park, PA 16802
(516) 640-8937

April 15, 2013

Dr. Glenn Palmer
Professor, Department of Political Science
208 Pond Lab
Pennsylvania State University
University Park, PA 16802

Dear Dr. Palmer:

I am writing to express my support for the project titled "Using crowdsourcing to measure complex social concepts: A pilot study." I believe this study has the potential to greatly impact the discipline of Political Science as well as to influence the many other social science disciplines that utilize content analysis to measure meaningful concepts of interest.

In the field of international relations, since researchers cannot be physically present to observe events as they are taking place, much of the data we collect come from second-hand sources such as news reports. Structuring the information in news reports into meaningful data is a time-consuming and repetitive task; it is also an essential part of the scientific process. The method being tested in this pilot study improves the efficiency as well as significantly reduces the redundancy associated with the coding process in the field of international relations.

The ideas put forth in this proposal utilize the strengths of both artificial and human intelligence for measuring complex social concepts. I have studied each of these approaches towards measurement, both in my home department (Political Science) and in the College of Information Sciences and Technology, where I have completed the coursework necessary for minoring in that discipline. The research design proposed here is novel and promising, and I look forward to working with you to further develop the method and conduct the analyses.

Sincerely,

Vito D'Orazio

Dr. David Reitter

PHONE
+1 814 867 3159

EMAIL
reitter@psu.edu

WEB
www.david-reitter.com

OFFICE
College of Information Sciences
Penn State University
316 IST Building
University Park, PA 16802
U.S.A.

To whom it may concern

April 3, 2013

In Support of the Project "Using crowdsourcing to measure complex social concepts: A pilot study."

Dear Professor Palmer,

I am writing this to express my support of your proposed project.

As you know, the availability of large data collections has transformed one of my core research fields, computational linguistics. It has made real-world applications such as speech recognition and machine translation possible.

More recently, we are seeing substantial, novel work in a related area, computational psycholinguistics, which has the potential to redefine research on human language processing, which has traditionally been the domain of psychology and low-volume, tightly-controlled, lab-based experimentation. Even experimental researchers rely on online platforms, chiefly Mechanical Turk, to generate data.

These methods come with caveats w.r.t. ensuring validity of data and experimental control. Big-data social science has to answer similar questions with observational and crowd-sourced data. Thus, I am confident that your proposed project will make great strides towards making crowd-sourced analysis useable by the political sciences.

As you know, the College of IST has the computational and intellectual resources to implement the machine-learning component and the Mechanical Turk interfaces necessary to facilitate the planned experiment. In anticipation of a joint proposal submitted to the NSF and a joint publication, my Lab will provide my time without charge, along with a Research Assistant who will implement the pre/post-processing A.I. and Mechanical Turk interfaces, as well as ad-hoc computational resources to host and process data if needed.

I am looking forward to working with you on this.

Dr. David Reitter
Assistant Professor of Information Science and Technology
Penn State University

PENNSTATE

Dr. Glenn Palmer
Professor
Department of Political Science
Pennsylvania State University
208 Pond Lab
University Park, PA 16802

Dr. Palmer:

I am writing to express my support for the project "Using crowdsourcing to measure complex social concepts: A pilot study." I am excited for the opportunity to be involved in this project, which I believe offers the potential to bridge several disciplines in developing an efficient and cost-effective way of analyzing media texts.

Like political science, communications relies upon trained coders — typically graduate or undergraduate students — to process media texts as a source of information about real-world events and the ways in which they are covered by the media. This process is time-consuming and ties up skilled individuals in repetitive tasks. Moreover, the repetitive nature of the process risks shortcomings in data due to fatigue. Finding an alternative approach is a promising concept.

As someone who has carried out numerous content analyses and contributed to the methodological literature on the topic, I believe the notion of crowd-sourcing at least certain types of coding has great promise. The Internet has proven to be a great source of distributed expertise, as demonstrated by projects such as Wikipedia and Project Guttenberg. Understanding how to use those resources for coding offers a chance to develop an effective methodology and also think carefully about the operationalization and validity of key variables currently measured through content analysis. I look forward to collaborating with you and the other collaborators on this promising project.

Sincerely,

Mike Schmierbach
Assistant Professor
Department of Media Studies
College of Communications
Pennsylvania State University

Biosketch for Glenn Palmer

Glenn Palmer (Ph.D., University of Michigan 1985) is professor of political science at Penn State University. His most recent research looks at the role of domestic politics on foreign policy as well as the effect of international conflict on domestic political and economic conditions. His work has appeared in a variety of journals, including *International Studies Quarterly*, *American Political Science Review*, *American Journal of Political Science* and the *Journal of Conflict Resolution*. His co-authored book, A Theory of Foreign Policy, was published by Princeton University Press in 2006. He is the executive director of the Peace Science society (International) and editor of *Conflict Management and Peace Science*. He has been the principal investigator of the Militarized Interstate Dispute Project since 2002 and has received three NSF grants supporting his research on this project since 2007.

Biosketch for Vito D'Orazio

Vito D'Orazio is a PhD candidate in the Department of Political Science at Penn State (PhD expected December, 2013). For the 2013 – 2014 academic year he has accepted a fellowship to work at Harvard University's Institute for Quantitative Social Science. From 2009 – 2011, he worked for Dr. Glenn Palmer on the NSF-funded Militarized Interstate Dispute project, primarily in applying document classification and named entity recognition methods to the content analysis of news reports. He has remained associated with the MID project since. For the 2011 – 2012 academic year, D'Orazio was a predoctoral fellow at Penn State's Quantitative Social Science Initiative where he studied applications of machine learning methods to Political Science. Beginning in the summer of 2012, he has worked for Dr. Joseph Wright on a project funded by the Office of Naval Research through the Minerva Research Initiative. He has also worked for Dr. Philip Schrodt on the DARPA-funded Integrated Crisis Early Warnings Systems project.

BIOGRAPHICAL SKETCH

NAME			POSITION TITLE
David T. Reitter			Assistant Professor

eRA COMMONS USER NAME (credential, e.g., agency login)			
DAVID.REITTER			

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
University of Potsdam, Potsdam, Germany	Dipl.	2003	Comp. Linguistics
National University of Ireland, Dublin, Ireland	MSc.	2005	Computer Science
University of Edinburgh, United Kingdom	Ph.D.	2008	Cognitive Science
Carnegie Mellon University, Pittsburgh, U.S.		2012	Cognitive Psychology

A.   PERSONAL STATEMENT

The proposed research investigates how a combination of crowd-sourcing and artificial intelligence can enhance data collection in the social sciences by obtaining annotated structured data in large quantities, improving its accuracy and timeliness.

Dr. Reitter's research has been at the intersection of computational linguistics and big-data cognitive science. In recent years, Reitter has investigated related questions, pointing out the interplay between individual-level cognition and network-level effects. Reitter has backgrounds in the fields relevant to these questions, specifically in Cognitive Science, Linguistics (discourse) and in Computer Science. He combines methods from these fields to study the relationship between such micro/macro-level mechanisms.   For example, he employed data-driven computational methods to test psycholinguistic hypotheses about adaptation in dialogue. The project used small- and large-scale cognitive simulation (e.g., with ACT-R) in combination with empirical experimentation to model learning and adaptation in human subjects, specifically during interaction within pairs and larger groups. Reitter's work uses cognitive modeling and network simulation techniques that have allowed a recent growth in interest for the interaction of cognitive mechanisms with the social environment.

Dr. Reitter is familiar with statistical modeling and machine learning techniques commonly employed in big-data computational linguistics research. His use of statistical and cognitive modeling frameworks, validated and refined through careful experimentation, as well as computational tools facilitates the simulation of human societies at a larger scale than what lab-based research allows.

B. POSITIONS AND HONORS

Academic Positions

2002–2004   Research Fellow, MIT Media Lab Europe, Dublin, Ireland

2008–2010   Postdoctoral Researcher, Department of Psychology, Carnegie Mellon University

2011–2012   Research Psychologist (Special Faculty), Department of Psychology, Carnegie Mellon University

2012–   Assistant Professor (tenure-track), College of Information Sciences and Technology, Penn State University

Honors

2003   Best Thesis Award,
        Society for Computational Linguistics and Language Technology 2003

2005   Doctoral Scholarship (3.5 years): Edinburgh-Stanford Link

2008   Postdoctoral Fellowship (2 years), German Academic Exchange Service (declined)

2009   Best Paper, Behavior Representation in Modeling and Simulation Conference

2009   Best Model, Predicting Cognitive Performance in Open-ended Dynamic Tasks,
        A Modeling Comparison Challenge, International Conference on Cognitive Modeling

Other Experience (selection)

2005/06   Board Member, Assoc. Computational Linguistics: SIG *Natural Language Generation*

2005–   Lead and founder, *Aquamacs*, a widely used application for science & engineering

2011/12   Workshop Co-Chair, 2nd and 3rd Workshops *Cognitive Modeling and*
        Computational Linguistics

2012   Editor, Issue on Computational Models in Linguistics,
        Topics in Cognitive Science

C. SELECTED PEER-REVIEWED PUBLICATIONS

Most relevant to this application

D.R. and Christian Lebiere.   How groups develop a specialized domain vocabulary:

A cognitive multi-agent model. *Cognitive Systems Research*, 12(2):175-185, 2011.

Kevin A. Gluck, Clayton T. Stanley, Jr. L. Richard Moore, D.R., and Marc Halbrügge. Exploration for understanding in cognitive modeling. *Journal of Artificial General Intelligence*, 2(2):88-107, 2010.

D.R.. Metacognition and multiple strategies in a cognitive model of online control. *Journal of Artificial General Intelligence*, 2(2):20-37, 2010.

D.R. and Christian Lebiere. Towards cognitive models of communication and group intelligence.
In *Proc. of the 33rd Annual Meeting of the Cognitive Science Society*, pp. 734-739, Boston, MA, July 2011.

D.R., Katia Sycara, Christian Lebiere, Yury Vinokurov, Antonio Juarez, and Michael Lewis.
How teams benefit from communication policies: information flow in human peer-to-peer networks.
In *Proc. 20th Behavior Representation in Modeling & Simulation (BRIMS)*, 8 p., 2011.

D.R., Ion Juvina, Andrea Stocco, and Christian Lebiere. Resistance is futile: Winning lemonade market share through metacognitive reasoning in a three-agent cooperative game. In *Proc. 19th Behavior Representation in Modeling & Simulation (BRIMS)*, 9 p., Charleston, SC, 2010.

D.R. and Christian Lebiere. Towards explaining the evolution of domain languages with cognitive simulation.
In *Proc. 9th International Conference on Cognitive Modeling (ICCM)*, 8 p., Manchester, UK, 2009.

Further publications relevant to the field

D.R., Frank Keller, and Johanna D. Moore. A computational cognitive model of syntactic priming.
Cognitive Science, 35(4):587-637, 2011.

D.R. and Christian Lebiere. A cognitive model of spatial path planning.
Computational and Mathematical Organization Theory, 16(3):220-245, 2010.

D.R. and Christian Lebiere. Accountable modeling in ACT-UP, a scalable, rapid-prototyping ACT-R implementation. In *Proc. 10th International Conference on Cognitive Modeling (ICCM)*, pp. 199-204, Philadelphia, PA, 2010.

D.R. and Christian Lebiere. A subsymbolic and visual model of spatial path planning.
In *Proc. Behavior Representation in Modeling and Simulation (BRIMS)*. 8 p., 2009.

D.R. and Frank Keller. Against sequence priming: Evidence from constituents and

Biosketch for Dr. Michael Schmierbach

Dr. Schmierbach (Ph.D., University of Wisconsin-Madison, 2004) is an assistant professor in the Department of Media Studies at Pennsylvania State University. His research explores the psychology of media effects and the development of effective methods for communication research. He has published more than two dozen papers in such journals as Communication Research, Journal of Communication, Media Psychology and Mass Communication and Society. Some of his recent projects include investigations of content analytical techniques for video games, the connection between game difficulty and enjoyment, credibility judgments related to Twitter, and the role of media-generated exemplars in creating third-person perceptions. He is currently under contract for a textbook on communication research methods, including content analysis, and serves as research chair for the Communication Theory and Methodology division of the Association for Education in Journalism and Mass Communication.

Vice President for Research	The Pennsylvania State University	Phone : (814) 865-1775
Office for Research Protections	The 330 Building, Suite 205	Fax: (814) 863-8699
		Email : orprotections@psu.edu
		Web : www.research.psu.edu/orp

Date: April 24, 2013

From: The Office for Research Protections - FWA#: FWA00001534
Stephanie L. Krout, Compliance Coordinator

To: Glenn Palmer

Re: Determination of Exemption

IRB Protocol ID: 42955

Follow-up Date: April 23, 2018

Title of Protocol: Using crowdsourcing to measure complex social concepts: A pilot study

The Office for Research Protections (ORP) has received and reviewed the above referenced eSubmission application. It has been determined that your research is exempt from IRB initial and ongoing review, as currently described in the application. You may begin your research. The category within the federal regulations under which your research is exempt is:

45 CFR 46.101(b)(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

Given that the IRB is not involved in the initial and ongoing review of this research, it is the investigator's responsibility to review IRB Policy III "Exempt Review Process and Determination" which outlines:
- What it means to be exempt and how determinations are made
- What changes to the research protocol are and are not required to be reported to the ORP
- Ongoing actions post-exemption determination including addressing problems and complaints, reporting closed research to the ORP and research audits
- What occurs at the time of follow-up

Please do not hesitate to contact the Office for Research Protections (ORP) if you have any questions or concerns. Thank you for your continued efforts in protecting human participants in research.

This correspondence should be maintained with your research records.