Cognition, Computation, and Human Data Science

Ramesh Balasubramaniam, Yihsu Chen, Rick Dale, Carolyn Jennings, Evan Heit, Chris Kello, Paul Maglio, Teenie Matlock, David Noelle, Michael Spivey, Peter Vanderschraaf, Anne Warlaumont, Tony Westerling, Art Woodward, and Jeff Yoshimi

Executive Summary of Initiative: Cognitive and Information Sciences (CIS) is flourishing at UC Merced, and worldwide as an interdisciplinary growth area for science and engineering. We propose to build on our track record of excellence in research and education. We plan to maintain the steady pace of growth in CIS that has built our cohesive unit, while also fostering numerous connections to other groups on campus and within the UC system. We also outline how these connections could be expanded through an initiative in *computation and data science* (see corresponding proposal), with an emphasis on human activity in its many forms, across many scales, and in its many environments. In this document, we provide rationale in the context of campus-wide goals and national trends, and outline plans and resource needs for realizing this vision.

Intellectual Components of the Initiative: The study of cognition is on the brink of discoveries that could be as profound as any in the history of science. From consciousness to artificial intelligence, from physical bases of thought to first principles of human intelligence and behavior, cognitive science research is burgeoning with new theories, methods, and technologies. Along with the boom in basic research, more and more information-based and learning-based applications are being incubated, developed, and implemented. Correspondingly, investments in cognition are being made by universities, funding agencies, and companies worldwide.

UC Merced has an opportunity truly unique to this time and place. Our new research university has proven to be fertile ground for Cognitive Science to take root. Our founding faculty had the pioneering vision and gumption to stake out a claim for the interdisciplinary study of brain, mind, technology and society. More faculty members have since come to join the cause, and a cohesive group has formed around language, dynamics, and computational modeling. Together we have written one of the campus' most successful storylines:

- The Cognitive Science undergraduate program recently underwent review and received uniform admiration. Reviewers praised the hands-on laboratory research opportunities for undergraduates, including our 151 current Majors whose demographics are well balanced across gender and ethnicity. Reviewers were impressed by how many were admitted to first-rate Ph.D. programs, such as Brown University, Indiana University, Purdue University, and UC Santa Cruz.
- CIS was the first Ph.D. program to be born out of the Interim Individualized Graduate Program, and is now the most selective on campus. In AY2012-13, our 21 students collectively had 40 papers submitted, accepted, or published; gave 59 conference presentations in 10 cities and 6 countries; and attended a dozen professional workshops and summer schools. Our first Ph.D. recipient recently published as first author in the Proceedings of the National Academy of Sciences, and another went straight to a tenure-track faculty position at the University of Memphis.
- This summer, the Cognitive Science Society's annual meeting was attended by over 1200 people from more than 300 institutions worldwide. There were more presentations from UC Merced than any other institution, including Stanford, Berkeley, and MIT. While these are aspirational peers, we are approaching Indiana University, University of Rochester, and UC San Diego—we are punching above our weight.
- Accolades of faculty delivering the CIS curriculum include extramural awards totaling more than $4M at UC Merced, at least eight editorships, a dozen or so leadership roles in professional societies, three current/former standing panel members at NSF or NIH, and 6 h-index impact factors of 19 or more.

UC Merced has an opportunity that is not just to build a world-renowned cognitive science program—we are doing that already. *Our campus can be a leader in making cognition the next big advance of science. Questions we work on—such as those surrounding consciousness, learning and memory, and human cooperation—are featured in top lists of outstanding scientific problems (e.g. Science, July, 2005). Big advances towards solving these problems would be on par with theories of evolution and quantum mechanics, and on par with the discoveries of DNA and the electron. Our campus can be a leader in training the next generation of scientists and engineers who translate basic advances in cognition into new businesses, technologies, and therapies.* The remainder of this document outlines one possible way we might realize this CIS vision, and connect it with broader themes on campus.

Definition of Thematic Area: Our proposal connects with computation and data science as overarching, interdisciplinary themes at UC Merced. These themes are emerging in a number of other groups, including the Sierra Nevada Research Institute, Center for Quantitative Research, Center for Computational Biology, Center for Human Adaptive Systems and Environments (CHASE), Applied Mathematics, Computer Science, Mechanical Engineering, Quantitative and Systems Biology, and Chemistry and Chemical Biology, and the proposed Center for Theory and Computation. Here we summarize computation from the CIS point of view, as it relates to the study of cognition.

Many past advances throughout the sciences were supported by analytical work in mathematics. More recently, the advent of computation has opened up a complementary, algorithmic means of formalizing theories and hypotheses. Indeed, cognition itself may be a form of computation yet to be determined, and formal theories of cognition are usefully expressed as computational models when their complexities lie beyond the reach of analytic solutions. We think that computational modeling will play a major role in making the next big advances in cognitive science, and the same could be said for many other areas of science and engineering.

Computational methods also are essential for analyzing large-scale datasets, especially those capturing the complexities of human activity, broadly construed. In fact we named our graduate program and bylaw unit "Cognitive and Information Sciences" to signify our emphases in computational modeling, methods, and related technologies. These emphases provide the common ground for a number of interdisciplinary connections based in cognitive science.

Faculty Participation: These bullets provide some examples of current faculty projects and collaborations that illustrate interdisciplinary connections between CIS and related areas of research represented in other proposals.

- Ramesh Balasubramaniam and Michael Spivey are collaborating with Stefano Carpin (EECS) to acquire funds for state-of-the-art laboratory equipment to study human cognition and problem solving using tools from robotics and non-invasive brain stimulators.
- Rick Dale has an NSF-funded project with Suzanne Sindi (Applied Math) and David Ardell (Biology) to seek common principles that guide the evolution of structures both at the molecular and linguistic levels. The project employs computational and mathematical modeling, along with human data as a test bed for exploring how selection guides the emergence and change of structures across scales.
- Chris Kello has collaborated with Stefano Carpin (EECS), Paul Maglio (Mgmt), and David Noelle on a DARPA-funded project with IBM to develop hardware, software, and test environments for next generation neural processing units. Kello and Carpin are also seeking funding to collaborate with YangQuan Chen (Mechanical Engineering) on human-robot search teams (also see proposal for the *California Institute of Drone Engineering Research*).
- Teenie Matlock, Founder and Director of the Center for Climate Communication (SNRI), works with UCM environmental scientists on how humans communicate about climate adaptability and risk. She works with engineering faculty on NSF-funded projects to develop avatars that can

communicate with humans in virtual environments as well as with Berkeley (ICSI) researchers on an IARPA-funded project to develop a large-scale multi-lingual metaphor repository system.

- Anne Warlaumont and Ajay Gopinathan (Physics) are seeking funding to apply techniques for analyzing and simulating animals' spatiotemporal foraging patterns to infants' foraging in acoustic information space. She is also collaborating with Eric Walle (Psychology) on a study of infant-adult interaction patterns in daylong audio recordings of families from the Central Valley, in order to better understand the relationship between language acquisition and locomotor development.

These projects demonstrate how CIS research and education goes beyond computational models of cognition per se. For us, computation means human data science. It means training students with the skills needed to compete for jobs in academia and industry, in a world where more and more data are readily available on human activities of all kinds, at all scales. From genetic to cellular to individuals, groups, and organizations, new methods and technologies are being developed to extract information from rich, heterogeneous datasets on human activities. This complex, multi-scale perspective on human activity lies at the heart of CHASE, which is elaborated in its own corresponding contribution to the strategic focusing initiative. *Here we focus on putting UC Merced at the forefront of theoretical advances in cognition, as well as methodological and technological advances in human data science.*

UCM's Role: The Chancellor has set a goal of 10,000 students by 2020, with 1,000 graduate students. We need to achieve this goal while maintaining quality, affordability, and accessibility for California residents, as well as expanding diversity whenever and wherever possible. The campus is off to a good start in many respects, but one of our biggest challenges will be reaching 1,000 graduate students. We also have one of the largest proportions of STEM programs and students not only in the UC system, but of research universities across the country. This proportion is consistent with the campus' originally envisioned STEM focus. We now have an opportunity to revisit our focus, and consider ways to enhance efforts to maintain affordability and expand diversity moving forward.

CIS is in an excellent position to contribute to campus-wide goals in the context of our current STEM focus. CIS is STEM-heavy, yet based in the School of Social Sciences, Humanities, and Arts. Being the most selective Ph.D. program on campus, CIS has a demonstrated capacity to increase the number of quality Ph.D. students given advising capacity—currently we have an average of 2.6 Ph.D students per faculty member, and our goal is to increase this average to about 3.0 at steady state. CIS also draws a large proportion of female students and researchers relative to many other STEM disciplines (about 50% of our current graduate student population is female).

Finally, within the year, we plan to submit a proposal for a Master's program in CIS focused on human-technology interaction and human data science, areas in which students will gain skills in high demand in Silicon Valley, as well as academia. The program would have natural connections to the new proposal for a *School of Innovation, Management, and Economics*. The program will be designed so that Cognitive Science majors who graduate from UC Merced can earn the MS in one year, depending on courses taken as an undergraduate. Students without the necessary coursework, e.g. those coming from other universities in the Central Valley or California, would be able to finish the program in two years. As part of the proposal process, we will draw upon recent market analyses and work with administrative leadership to determine whether the program should be self-supported, or whether a model could be formulated in which our MS and Ph.D. programs mutually support each other. There are also plans to create an honors track for the Cognitive Science undergraduate Major, to formally recognize the outstanding undergraduate students already conducting honors-level research in our labs. In addition, philosophy and cognitive science are highly integrated at UC Merced, and plans are being developed to expand the CIS graduate program to include additional training for students studying the philosophy of cognitive science (see *Applied Philosophy* proposal).

Special Programmatic Needs: One of the primary purposes of strategic refocusing is to set a course for our academic mission, so that space and faculty resources may be directed accordingly. Thus far, the CIS

undergraduate and graduate programs have grown by one faculty member per year, with an average of about 500 sq ft of dry lab space per faculty member. We have focused on language, dynamics, and computational modeling as our brand, which has created a tightly integrated set of research programs with international renown. To complement, our connections with other programs have grown organically, without explicit plans or mechanisms to coordinate hires or space.

Looking forward, we believe that achieving our vision will be best served by fostering two modes of growth. One mode is to continue growing the core CIS group by one faculty hire per year, with commensurate student growth, including the addition of a CIS Master's program. This core mode will help maintain the cohesion that has been critical to our success. The other mode is to coordinate hires more explicitly with other groups and programs, and to strengthen interdisciplinary research and education in human-technology interaction and data science. Both modes of growth are exemplified in the FTE request submitted last year by the CIS graduate group. Four faculty hires were proposed in each of the following areas: Computational linguistics, network science, science communication, and computational systems neuroscience. The computational linguistics search is underway, and the network science position is planned for next year. Network analyses can provide insights across many kinds of human activities, from neural to semantic to social, and connections with multiple groups on campus.

To coordinate more explicitly with other programs and groups on campus, we offer for consideration *an initiative in human data science, perhaps as part of a larger computation and data science initiative*. The time is ripe for data science, as evidenced by a number of recent investments by institutions and agencies. For instance, the University of Rochester recently announced a $50M initiative, coordinated by their Institute for Data Science, to hire 20 new faculty "who are excited about engaging in collaborative research that connects advances in computational models and methods to other fields of engineering or life, social, or physical sciences" (http://www.rochester.edu/rocdata/recruit/interdisciplinary). Data science is the center piece of their 5-year strategic plan, and the hiring initiative seeks to make primary and secondary appointments across programs and disciplines. As another example, NYU is teaming up with Berkeley and the University of Washington on a five-year effort to help researchers take advantage of ever-increasing amounts of information by improving both data management and data analysis (http://www.moore.org/newsroom/press-releases). The project, slated to cost nearly $40 million, is being underwritten by the Gordon and Betty Moore Foundation and Alfred P. Sloan Foundation. Finally, the Obama administration is investing $100M in a human brain initiative that relies heavily on interdisciplinary data science (http://www.whitehouse.gov/share/brain-initiative). There is natural overlap between the emerging focus area on *Human Health* (see corresponding proposal), and human data science. This overlap has great potential for fostering interdisciplinary teams to pursue big funding opportunities like the human brain initiative.

We believe that UC Merced has a unique opportunity to emphasize "big data" generated from human-environment systems, broadly construed (see CHASE). Data science is an emerging theme in many groups on campus, and across the academic spectrum, from digital humanities to mountain ecosystems. Moreover, research relevant to humans and their environments can be found in each of our three schools, and each of our graduate programs, in one form or another. The following resources would help support such a campus-wide initiative, as well as growth and research in CIS:

- **Computation and Data Science Cluster**. A computation and data science initiative will draw faculty members who advise Ph.D. students needing space and technology resources for computational projects. A large, shared space for computational projects and students across programs would foster interdisciplinary interactions, and be more efficient compared with individual, non-shared spaces. A shared space could include server room(s), student "bullpens", meeting rooms, and breakout areas, for instance. The space could include human data science as well as other computational groups and projects on campus with similar needs and goals.
- **Interdisciplinary TAships**. Interdisciplinary graduate programs may not have simple alignments with undergraduate programs or even schools, yet TAships need to be reliably available to students in all graduate programs, for training as well as support. To eliminate potential barriers to

interdisciplinary initiatives like human data science, TAships should be reliably available to all graduate programs based on faculty contributions to undergraduate and Master's education.

- **Multimodal Human Activity Lab**. Human neural and behavioral activity is richly multimodal and multiscale, and new methods and technologies are becoming available to study this richness in unprecedented detail. CHASE is planning a common lab for studying the real-time unfolding of human neural and behavioral activity, in individual and group settings. The planned tools include 3D motion-capture, robotic effectors, and non-invasive neuroimaging tools such as high-density EEG and transcranial magnetic stimulation (TMS). These tools will be used for studying and modeling data from complex brain dynamics using tools from network science. The lab would provide an interdisciplinary environment that is not possible in individual labs/program communities. For example, cognitive and social scientists could work with computer scientists, roboticists and health-related professionals on behavioral analyses with e.g. diagnostic or training applications. Neuroscientists and cognitive scientists could work with physicists and statisticians to relate computation with behavioral and brain dynamics. An NSF Major Research Instrumentation proposal is being written to fund much of the equipment needed to start such a lab.