# Parallel Processing

## Michael Wilde, a fellow at the Computation Institute and software architect at Argonne National Laboratory, works to develop systems to enhance scientific productivity in neuroscience and other domains. Wilde's work supports parallel processing and high performance computing for efficiency in data analysis.

Disciplines like neuroscience, genomics, and biochemistry employ complex instrumentation and measurement, and produce large quantities of data. Consequently, they require greater computing power to analyze that data and produce scientific insight.

"As computers become more powerful, they have also become harder to use. Computing power has never come for free. It has always required a greater and greater level of programming and technical sophistication among the user base," explains Wilde. Wilde's work allows scientific research to reach new heights of efficiency and performance and utilize state-of-the art computational resources.

The Computation Institute at the University of Chicago supports and fosters an environment that facilitates multidisciplinary research centered on leveraging computational power.

"To make progress in disciplines that need computation, individuals with both domain and computer expertise need to work together. On the computing side, we often have people very fluent in algorithms working with people very fluent at the systems level. With expertise from scientists working in specific domains, we are now able to craft solutions to problems we might not have known even existed," explains Wilde.

At the Computation Institute, Wilde works with Ian Foster, Ph.D., and colleagues in a group known as the Distributed Systems Laboratory, linking research and applications work from Argonne National Laboratory's Math and Computer Science Division and the University of Chicago. Specifically, Wilde works with programming paradigms to help scientists utilize parallel and distributed processing for more efficient data analysis.

*Parallel and Distributed Computer Processing*

Tasks with vast quantities of data require considerable computing power. Greater data volume leads to increased systems level complexity, because computing in parallel is harder than computing serially, and computing remotely is harder than computing locally.

For example, if the Library of Congress wanted to digitize and determine the reading level of all the books in the collection, a computer would first need to scan all the books and place the text into a useable digital form. As volumes of books were scanned, computer programs could turn that digitized data into useful information: programs could convert the scanned images to text, and programs could analyze the reading level of each volume. This alone would require a great deal of computing power, but the scale of the project (a library of 33 million volumes) would mean that many machines would need to work together to perform computations on terabytes and petabytes of data in a reasonable time frame.

"If the latest laptop computer has half of a terabyte of storage, you could imagine that this task would involve thousands and thousands of laptops worth of data. Not only would you need to scan and organize the individual files, you would need to group the files into structured sets, and place meta-data on those sets to associate them with other relevant information –the title and author of the book, what kind of book it is, how it was scanned, at what resolution it was scanned – and send this data to computing sites that have appropriate power to analyze your data," explains Wilde.

With parallel and distributed computing, there are many additional challenges, including scheduling the needed computer resources, translating the specific analysis program data into terminology that the remote distributed parallel systems can use, and flow control: the problem of taking millions of volumes and thousands of pages and then running them through this program text. Finally, programs may need to correct the resulting scans and provide quality assurance, and associate the extracted data with the original data.

In cognitive and social neuroscience, instead of pages of books, data comes from people or animals. A study using fMRI techniques produces digitized image data; anatomical scans as well as functional scans of each participant during different experimental conditions.

"These studies result in multi-terabyte datasets that must be run through programs that take many, many hours of processing for each participant and for each scan. Then, these datasets go through analysis phases where scientists use different statistical tools and methods to try to extract information from the digitized signals of the scan and test statistical hypotheses on the data. Each of these phases requires long programs run on many datasets. Sending each dataset through a chain of programs requires a lot of manual coordination or detailed, difficult programming," explains Wilde.

This programming is largely unrelated to the science of the experiment being performed; it is not code written to perform the statistical analyses on the data, or to program the prompts a participant sees during an experiment, but the coordination of computers required to complete the analyses.

Parallel computing, now prevalent in computer architecture, allows several computer processors to work on the same problem at the same time, conserving time and energy. Parallel computation configurations include many computers doing different tasks (multiple program-multiple data) and many machines performing the same tasks (multiple program-single data). However, in order to complete these tasks with maximum efficiency, specialized tools and scripting languages are required.

*Globus and Swift*

Wilde's initial work with this field began with the Globus Project, started in 1994 by Ian Foster, with collaborators at the University of Chicago, Argonne, and the University of Southern California. The Globus Project was started to foster international scientific collaboration and facilitate more efficient grid computing systems.

"Globus has played and continues to play a leading role in helping people use distributed computers. It tries to make a diverse set of machines in different locations work together seamlessly to allow a scientist to do more computation with greater ease," explains Wilde.

Globus took the first step in trying to tame the complexity presented by diverse machines and interfaces by providing a uniform operation layer to allow users with different hardware and software to maximize computing power, allowing these machines to work together. Globus provides a uniform authorization and authentication layer, providing remote users a single digital identity that allows them to access a computer grid consistent with the availability of the resources.

Wilde's focus today is the Swift scripting language. Like Globus, Swift is "middleware," a tool above the operating system but below the software or application layer. This operating layer forges the connection between the operating system and the applications, functioning as a higher-level operating system for end users.

"Distributed parallel machines are, by nature, harder to use, because the programmer is now responsible for coordinating many activities at once, and synchronizing the work of many computers. Swift attempts to hide this parallelism from scientists, and explain this remoteness," says Wilde.

Swift helps users run many copies of an application program across different datasets, and fans data out to resources that it has dynamically located and assigned. It specifies the coordination of running programs on datasets, similar to other scripting languages like Perl, Python, and Ruby, but is implicitly parallel. In Swift, the user specifies what operations need to run on a dataset, and Swift determines, based on dependencies in the user's specification, what can run, in what order, and with what degree of parallelism. It hides the complexities of finding remote resources, using Globus to locate them and run their operations.

"Swift also connotes speed, of course. As it happens, a swift is also a small bird, lightweight but extremely fast, that flies around in flocks, often in parallel formations, which makes it a good mascot for the Swift parallel scripting language," offers Wilde.

*Transparency and Data Provenance*

"A scientist need not be concerned with whether these types of programs run on one computer or many, or in a grid or in a cloud. This level of detail takes the scientist further away from the area of her expertise. Still, the goal for Swift is to make parallelism transparent," explains Wilde.

Swift tries to shield the end user from the details of distributed computing. Wilde points out that parallelism, distribution, and heterogeneity of systems should all be transparent to the user, and the reproducibility of results derived from such computations should be assured.

Use of these methods also supports data tracking capabilities. In the scientific research process, a new dataset is consistently derived from previous data, frequently through the application of a program. Because Swift has taken a methodical approach to that process of applying code to data, it produces an audit trail. These methods allow researchers to know what code was executed on which dataset and in what environment. In this way, the researcher has a high fidelity digital companion to their scientific logbook.

Wilde says, "We are actively working on ways to make querying that data more practical to the scientist, so that they can ask reasonable questions about how the results were derived. When scientists are about to publish a study, they want to be able ask, 'Did these three participants' data run through the same software stack or a different stack as these other five participants? Was there any discrepancy among the parameters used, the calibration factors, the code versions, and the computing environments that could contribute to any differences, noise, or other aspects of our conclusions?' Data provenance comes as a gift when you structure your computations in a way that abstracts the physical location of the computing resources."

*Future Directions*

Beyond continually scaling these tools and techniques to meet the data intensive demands of contemporary science, mastering the engineering of software and hardware development, and promoting implicit, transparent parallelism, Wilde would like to bring this computing power to disciplines beyond those currently using the systems.

Explains Wilde, "Our work strives to allow this newly available computing power to serve more disciplines than the initial targets of those kinds of platforms, the very numerically oriented, computationally intensive sciences. We would like to broaden the number of disciplines that can make use of this new power, and make scientific computing easier for all." •

# CCSN Member Profile: Susan Levine

## Members of the CCSN conduct research related to cognitive and social neuroscience in their field of expertise and with unique scholarly perspectives. Susan Levine, Ph.D., Chairman of the Department of Psychology and the co-director of the Center for Early Childhood Research at the University of Chicago, was interviewed about her work in early childhood cognitive development and the Spatial Intelligence and Learning Center.

*What is the focus of your work?*

With Professors Susan Goldin-Meadow and Janellen Huttenlocher, as part of the National Institute of Child Health and Human Development Program Project *Environmental & Biological Variation and Language Growth,* I conduct language development research. Our overarching research goal is to understand how language interactions between children and parents impact the child's language, literacy, and conceptual development over time. For this research, we visit participating family's homes every four months, beginning when children are fourteen months old. During these visits, we ask the primary caregiver to simply do what they normally do – we do not give them any special toys to play with or explicit tasks to perform. We videotape the parent-child interactions for ninety minutes, focusing on the child if the parent happens to leave the room. When we return to the lab, we transcribe all the language and co-speech gesture that was produced by the child and the parent. We code the utterances in various ways, depending on our particular research questions.

For example, one such question concerns parents' use of number words, and the particular ways in which they talk about number. Elizabeth Gunderson, a graduate student at the University of Chicago, and I found that parents varied widely in their frequency of use of number words. Children who heard more "number talk" during the five visits between fourteen- and thirty- months of age had a more developed understanding of cardinal number, a central mathematical concept. We also found that certain kinds of number talk were most predictive of children's cardinal number knowledge – in particular, parent number talk referring to object sets rather than just rote counting or labeling numerals, and parent number talk extending beyond set sizes of three. Such parent-child interactions are dynamic, as the child's actions and interests may influence parent behavior, and vice versa.

Many researchers want to determine the activities in which children engage (i.e., reading, counting, puzzle-solving), and at what age those behaviors emerge, and then see how those activities predict academic achievement or development. However, this is a complicated story to tell. For example, parents and children often read books together, but the individual who initiates the practice changes over time — when a child is fourteen months old, a parent might always initiate the activity, but when she is four years old, a child is likely to ask to her parent to read to her. We are striving to answer the questions, "What is the parent saying? How is it influenced by the child's behavior? And do these patterns change when parents initiate certain kinds of interactions?”

Through this work, we hope to identify the kinds of interactions associated with strong language and conceptual development. Then, to follow up on these findings from our naturalistic study, we hope to test whether this is the case in experimental studies.

*What is the Spatial Intelligence and Learning Center (SILC)?*

The Spatial Intelligence and Learning Center (SILC) is a National Science Foundation Science of Learning Center. The Center involves faculty and graduate students from Temple University, Northwestern University, the University of Chicago, the University of Pennsylvania, and the Chicago Public Schools. We have several related initiatives: to understand more about spatial cognition through basic research, to study the tools that we think could improve spatial learning, to use these tools in training studies to see whether we can improve spatial learning, and to translate these findings into real world learning environments, such as schools and museums. This work includes studies of the early development of spatial thinking (the core of my research) as well as studies examining spatial learning in older students, particularly those studying science and math (e.g., physics, organic chemistry).

The basic premise of the Center is that spatial learning is important, but underemphasized in educational settings. There is not a subject in school called ‘Spatial Learning’. As a society, we tend to believe that some people are spatial thinkers, and others are not. However, research supports the fact that spatial ability is malleable, and can improve through instruction, particularly instruction that makes use of strong learning tools such as analogy, spatial language, embodiment and gesture. Spatial ability is predictive of entry into the STEM disciplines (science, technology, engineering, and math), even when you control for other abilities, like math and verbal skills. People who have better spatial visualization skills are more likely to take STEM courses and work in STEM-related fields. Research has also found sex differences in spatial skills, particularly in mental rotation and navigation tasks, and we would like to determine the basis of those sex differences. Because spatial thinking is malleable, and an important predictor of achievement in the STEM disciplines at school and in the workplace, this research has the potential to increase the representation of females in these disciplines.

*What processes contribute to spatial thinking?*

While my research focuses on the cognitive processes that underlie mathematical and spatial thinking, social and emotional factors also contribute. My research with Sian Beilock, an associate professor at the University of Chicago, as well as Gerardo Ramirez and Eli zabeth Gunderson, graduate students at the University of Chicago, has shown that an individual can have perfectly good spatial thinking skills, but do poorly on spatial tasks because of anxiety stemming from societal stereotypes passed from teacher to student or parent to child, not unlike math anxiety.

Understanding spatial skill development requires both naturalistic (e.g., home observation) and controlled experiments (in the laboratory). For example, during home visits, we observe which children engage in puzzle play and for how long. In one phase of study, we found that approximately half of our sample played with a puzzle in our presence. Later, we found that the children who played with puzzles at home were better at mental rotation tasks in the laboratory at four and a half years of age. However, this data is correlational. We are now carrying out an experiment to determine if there is a causal relationship between this kind of play and spatial ability. Ultimately, our goal is to translate our findings into curricula and teacher practice. •

# Affiliate Lab Interview

## Tatia Lee

Tatia Lee, Ph.D., an affiliate of the Center for Cognitive and Social Neuroscience at The University of Chicago, and the leader of the Laboratory of Neuropsychology and Laboratory of Cognitive Affective Neuroscience at The University of Hong Kong, studies the relationship between psychological processes of behavior and human brain. On October 3, 2011, Lee and colleagues held the *International Symposium on Applied Neuroscience and Neuropsychology* in Hong Kong.

*What is the aim of the symposium?*

The International Symposium on Applied Neuroscience and Neuropsychology (ISANN), founded by the Laboratory of Neuropsychology and Faculty of Social Sciences of The University of Hong Kong, is a biennial event that aims to enhance understanding and promote knowledge exchange about brain and behavior relationships – the neural underpinnings of the many social cognitive affective processes happening in the brain that determine our behavioral presentation at every moment of our lives.

In line with the globalization of the University and the Faculty, since the inaugural symposium in 2007, our events have brought scientists from renowned research laboratories and institutes from all over the world to Hong Kong. Research networks and clinical exchanges between Hong Kong and these laboratories and institutes have been fostered. We are also committed to nurturing future scientists, and for this reason we have created opportunities that encourage dialogues between senior secondary students and scientists.

Following the success of our previous symposia, this year's symposium titled *International Symposium on Applied Neuroscience and Neuropsychology 2011: Surfing the Socio-Affective Brain* brought new and timely discoveries of brain processes underlying social behaviors and emotional presentations – interpersonal sensitivity, perceived social isolation, emotional consequence of social comparison, and the development of psychopathic traits. It was our greatest honor to have renowned scientists from Canada (Jorge Armony), Israel (Simone Shamay-Tsoory), and the United States (James Blair, John Cacioppo, Jean Decety) speaking in this year's symposium. We were also very glad to have three HKU research graduates co-host a Forum titled *Affective World*.

*What was the theme of this year's symposium?*

How the brain operates to execute socio-affective regulation is the theme of this year's symposium. Socio-affective functioning defines the quality of life of an individual. Our speakers lead the journey exploring the brain mechanisms underlying interpersonal sensitivity, perceived social isolation, emotional consequence of social comparison, and the development of psychopathic traits. Knowledge about our socio-affective brain will shed light on ways that promote mental health and social harmony.

*What is the focus of your research?*

Neuropsychology studies the human brain and its relationships with specific cognitive and affective psychological processes and behaviors. Marrying the strength of neuroimaging and behavioral methodologies, my team has focused our research on unraveling the neural correlates and processes of the human brain, especially those involving the prefrontal regions. Cognitive-affective prefrontal processes are the very processes that define human nature and individual characteristics. With a background in both clinical psychology and neuropsychology, I am privileged to have the opportunity to conduct research along the empirical-application spectrum, covering the normal and the pathological changes of brain processes across the life span. There are three main lines in my research: prefrontal cognitive- affective processes, neuroplasticity, and evidence-based assessment and intervention.

*What are prefrontal cognitive-affective processes, and how are they studied?*

Prefrontal cognitive-affective processes are many, and they include regulation, social cognition, deception, and risk-taking decision making. Cognitive and affective regulation enables an individual to control for impulses that counter the requirement for achieving adaptive behavioral goals. Indeed, appropriate self-regulation promotes learning and effective processing of information captured from the social context and environment. The execution of goal-relevant behaviors requires a brain system that is capable of inhibiting habitual responses to orchestrate behavioral outputs. Achieving such a goal of response regulation is the foundation of intelligent behavior, and having efficient and effective cognitive and affective regulation systems are the prerequisites. We have identified the neural correlates of the cognitive and affective regulatory systems. Violence and addiction behaviors relating to inadequate cognitive-affective regulation have been studied. From the findings, a prefrontal hypoactivity accompanied by task-specific neural hyperactivity model to explain and understand impulse control disorders has been proposed. Further experiments have been planned to verify the validity of this model. Other studies have been launched to understand the cognitive-affective processes of the human brain.

Social cognition allows an individual to keep pace with the ever-changing social context in order to produce socially appropriate behaviors. Social cognition is the cognitive processes underpinning satisfactory and rewarding social interaction and relations. The areas studied are mentalizing (the awareness of the mental states of self and others), emotion recognition (the decoding of emotional cues delivered via facial expression and prosody), and empathy [Ho, N.S.P. and Lee, T.M.C., manuscript under preparation]. Understanding the neuropsychological foundation of social cognition will shed light on intervention for promoting prosocial behaviors, which is essential to social harmony as well as successful social integration of people suffering from various clinical conditions.

Deception and risk-taking decision making determine the quality of behavioral output of an individual. Sometimes, we need to withhold the truth, whether for malicious or benign reasons. The involvement of brain processes in the manipulation of information suggests that brain activity during deception and truth telling should be different. It is this very assumption that founds the theoretical basis of neuroimaging studies on deception. My team was among the first to apply functional magnetic resonance imaging (fMRI) methodology to deception research. Using different experimental tasks in studies of deception, we identified that activity in the prefrontal, cingulate, and parietal regions is associated with lying (increased brain activity during lying, relative to a truth-telling control). To fill a significant gap in fMRI deception research, my team has published one of the first papers on the impact of emotional attributes of suppressed information on deceptive responses. We have expanded our work to look at the specificity and generalizability of ours and others' findings.

Risk-taking decision making depends on a series of cognitive and affective processes that aim to balance the potential losses and benefits of an action. We have studied the effect of gender on neural activity during risk taking, as well as other factors and intervention related to risk taking. Women, relative to men, are more conservative and show more emotional arousal during risk-taking. Also, our data indicate that aging seems to be associated with decreased risk-taking.

*What is neuroplasticity?*

The human brain has the capacity for plastic reorganization through learning and experience, sometimes called unimodal neuroplasticity. Training could induce plasticity presented as experience-dependent dendritic sprouting, which guides neurons to connect with each other after learning. Animal and human research findings of my team and that of the others have clearly revealed that enriched experience offered by rehabilitation can enhance reorganization in the adjacent intact cortex and contribute to functional recovery after strokes and other brain trauma.

Likewise, multiple neural correlates may be involved in a single action, or cross-modal neuroplasticity. Each neural correlate may be accounted for in multiple functions. This complex property of our brain leaves us the possibility to affect different functions indirectly. In fact, these kinds of cross-modal stimulation have been commonly adopted in neuro-rehabilitation. In this context, we have examined cross-modal neuroplasticity in the temporal and occipital regions and in the prefrontal regions.

We are also interested in the proliferation of endogenous stem cells. Recently, more evidence supports the hypothesis that exercise could stimulate hippocampal neurogenesis, and this generation of new cells would lead to structural change in the hippocampal and frontal regions, thus leading to functional cognitive change. In this connection, we have identified that restored neurogenesis in stressed animals was associated with improved hippocampal-dependent memory and decreased depression-like behavior. Combining ours and others' work, evidence showing a positive relationship between exercising and neurogenesis is strong in animal models.

Combining the knowledge gained in the above three areas of neuroplasticity, my team endeavors to develop interventions effective for promoting brain health and maximizing the cognitive-affective recovery in people suffering from brain trauma.

*How does your research employ evidence-based assessment and intervention?*

Evidence-based clinical practice starts with the use of validated tests for assessment. In order to achieve this goal, much research effort has been devoted to study the psychometric properties of the tests so as to evaluate their applicability to the Chinese. My laboratory has been developing new and valid assessment procedures applicable to the Chinese. To consolidate our efforts in test development, a normative database of neuropsychological tests applicable to the Chinese population was published in 2003. This book has become an essential reference for research and quality evidence-based practice in the field of clinical psychology and neuropsychology. This line of work has been extended to the mainland, leading to the publication of the second edition (2009), and its revised version (2010), of the normative database applicable to Chinese people in Hong Kong and on the mainland, as well as other normative data. My team has been and will continue to provide consultation on cross-cultural neuropsychological practice to promote quality neuropsychological research and practice in other cultural and geographical settings.

Lastly, we are also interested in cognitive rehabilitation. To facilitate effective cognitive rehabilitation, my team has studied predictors of functional outcomes and the disease adjustment processes. We are actively conducting clinical trials for testing the effectiveness of clinical intervention strategies, another main foci of research of my lab. •