Framing CCC CSLS WS3

Kevin Sullivan, Chuck Friedman, Beth Mynatt
University of Virginia
Jan 24, 2017

Socio-Technical Systems

- Systems comprising people, machines, computing, physical elements
- Many service system examples: air-traffic control, healthcare delivery, ed, etc.
- Limited learning at systems scale, often glacial pace of system improvement

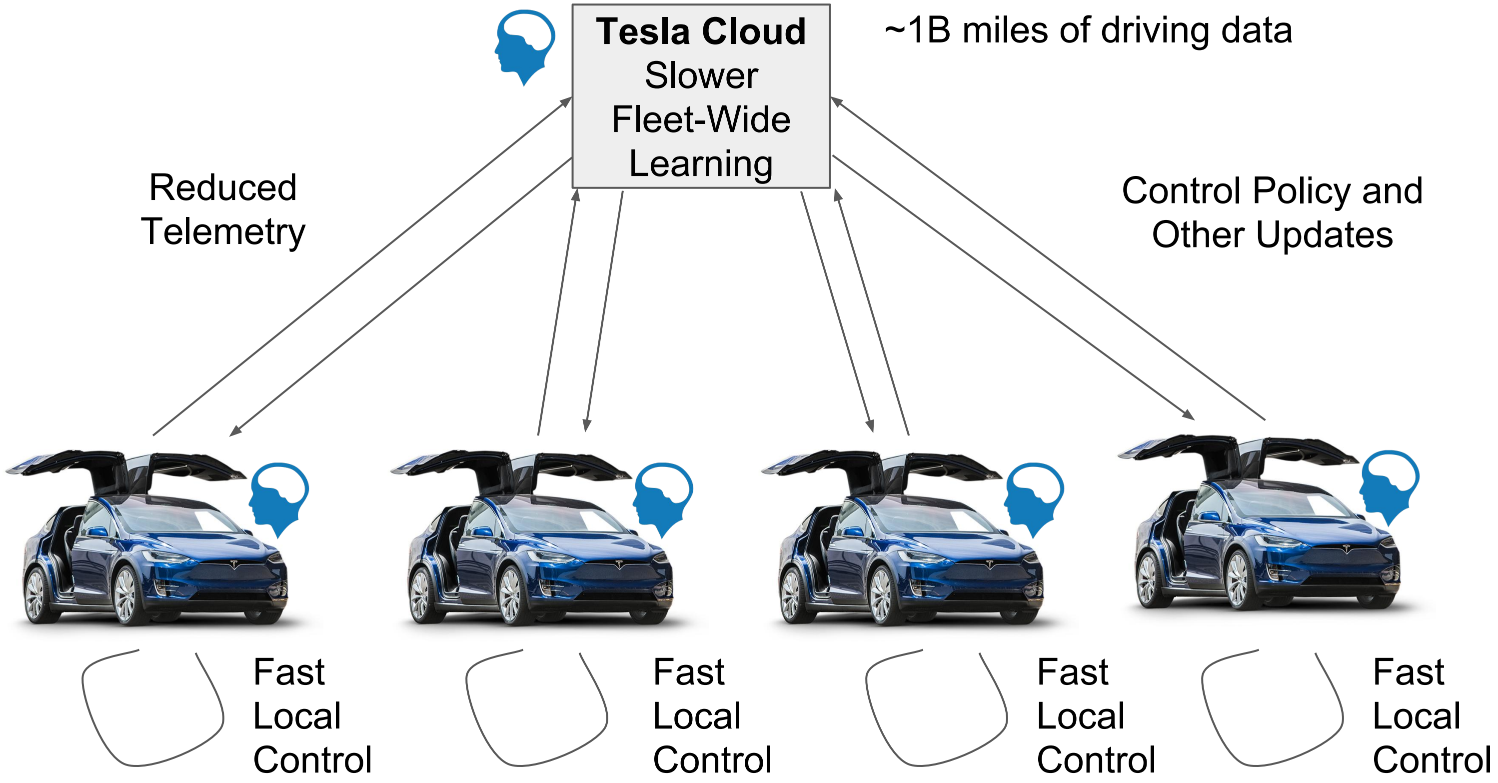
Human-in-the-Loop Cyber-Physical Systems

- Central objective is to optimize performance of physical devices/phenomena
- Enabled by advances in sensing and the integration of discrete, logic-based computing with continuous control via hybrid discrete-logic/physics models
- Increasingly use machine learning, and integration of logic and probability
- Often involve humans-in-the-loop to handle control when machine can’t
- Tendency is to progressively reduce human role, to the extent achievable
- Examples: Robotic manufacturing; learning fleets of autonomous vehicles

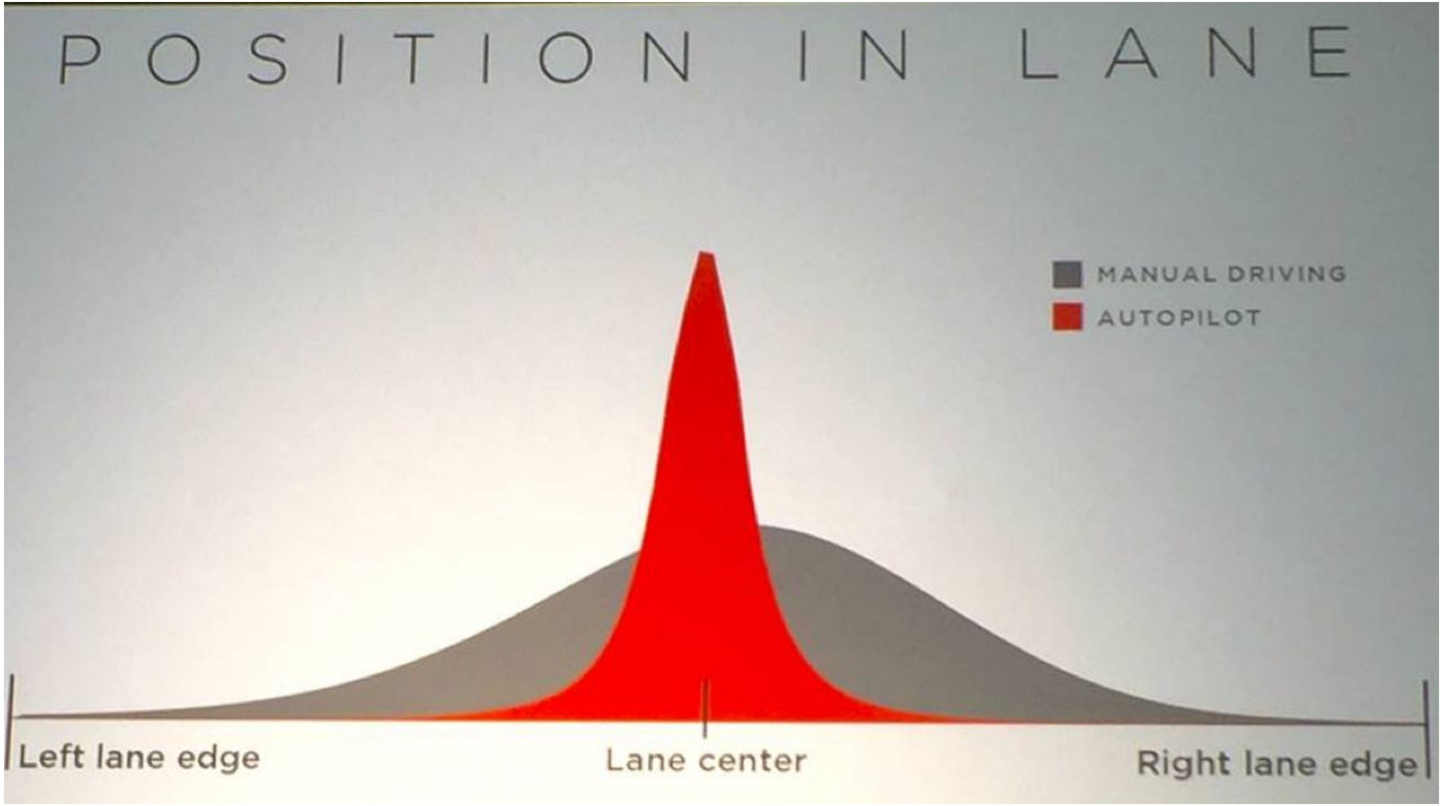
Example CPS: Mercedes Benz Manufacturing Line

Example of Human-in-the-Loop Learning CPS

Cyber-Physical Systems Learning Loop



Tesla - Performance Improvement (lane centering)



https://electrek.co/2016/05/24/tesla-autopilot-miles-data/

Tesla -- Performance Improvement (auto-steering)

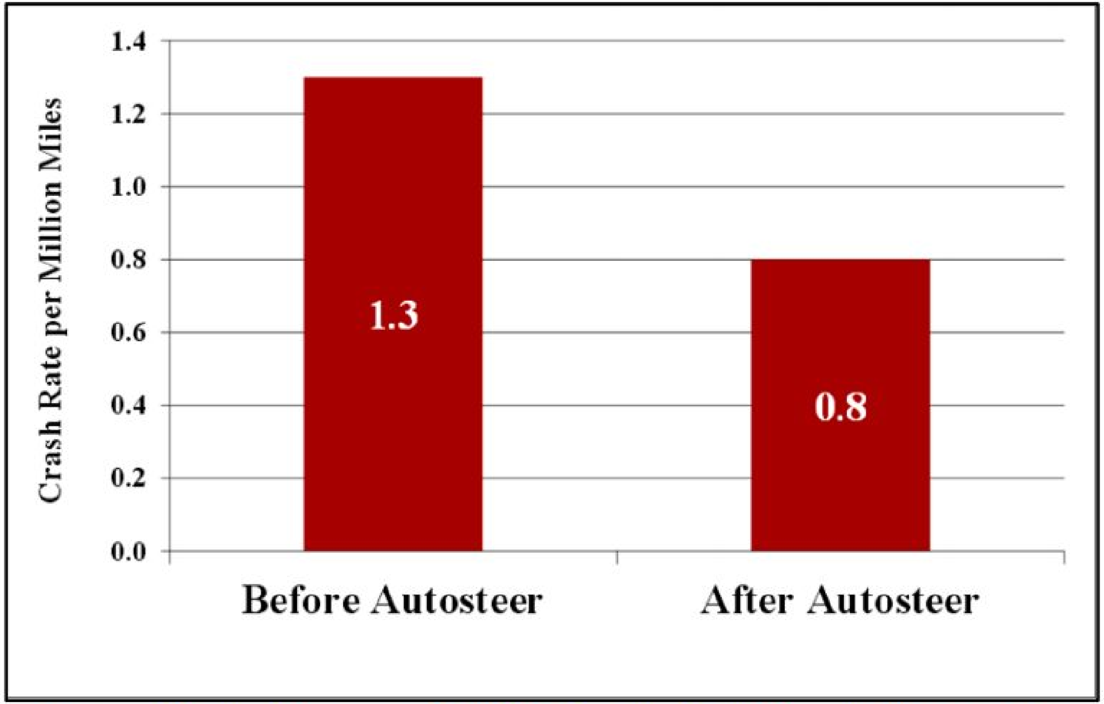
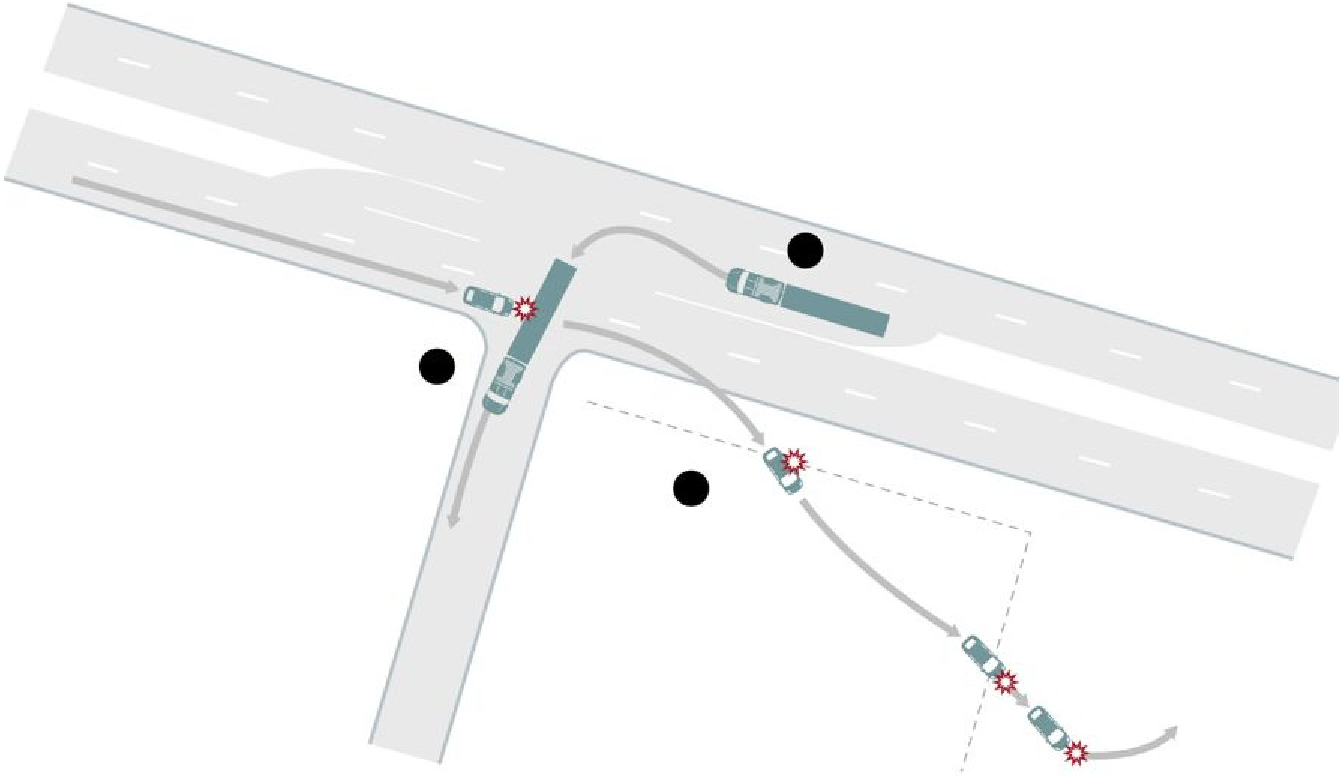


Figure 11. Crash Rates in MY 2014-16 Tesla Model S and 2016 Model X vehicles Before and After Autosteer Installation.

https://electrek.co/2017/01/19/tesla-crash-rate-autopilot-nhtsa/

Tesla Accident, Failure of Human-in-Loop Control

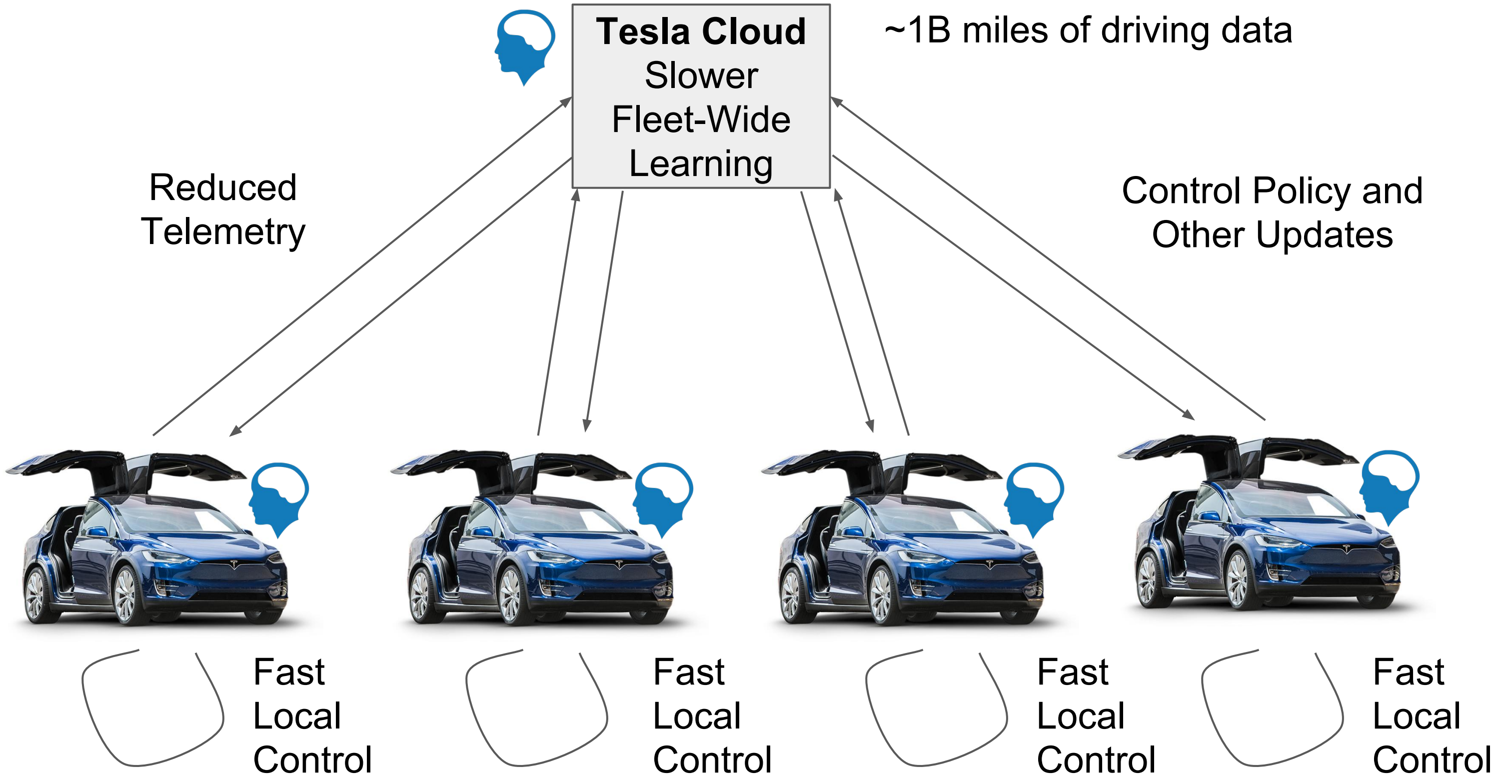
What happened



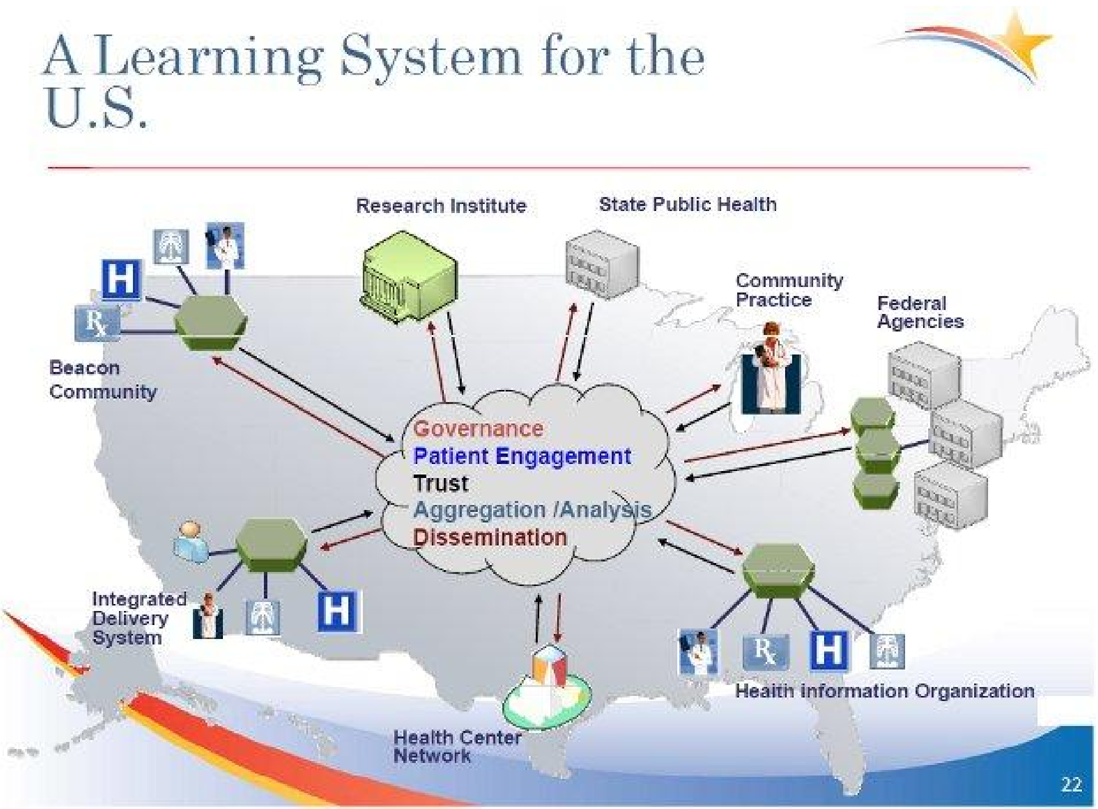
Trust and Reliance are Human Issues

- A second, less widely recognized failure mode in this case
  - Car continued autonomous driving after its “shearing” until it hit telephone pole
  - Just evaluating control functions on input sensor data and acting accordingly
  - No awareness whether operating within their envelopes of competence
- Issues of trust in, trustworthiness of, and reliance on AI/autonomy
- Mr. Brown over-trusted the technology relative to its actual capabilities
- Under-trust can be just as harmful; correct calibration of trust is required
- CPS inevitably edge into more complex human/social issues, but CPS foundational theories not designed to deal with such phenomena

Cyber-Physical Systems Learning Loop



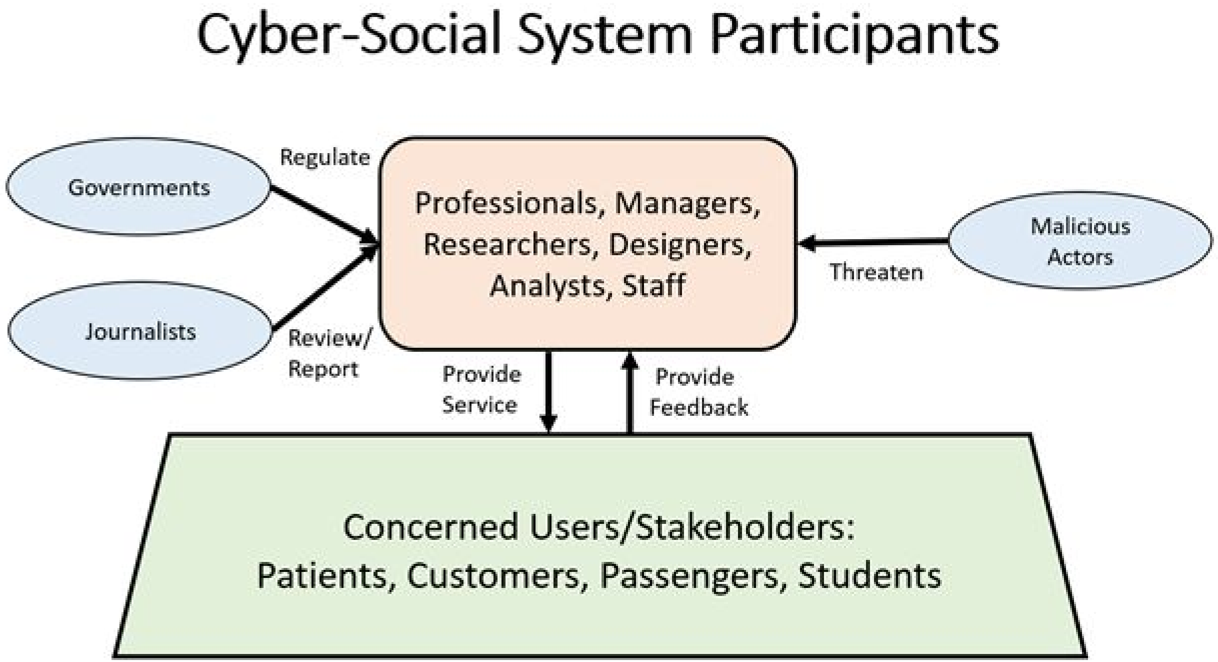
Cyber-Social Learning System for Health (LHS)



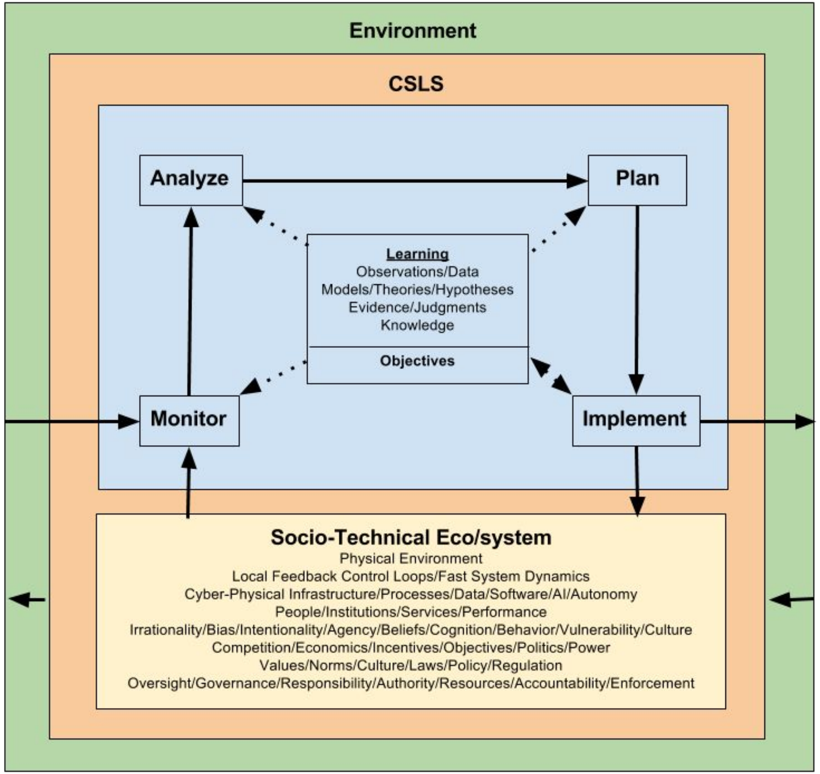
Cyber-Social Learning Systems

- Objective is to optimize performance of inherently human-intensive systems
  - Public services, healthcare delivery. education, intelligence, defense, food/energy/water
- Enabled by advances in and integration of CISE, Engineering, SBE, design, policy & law, implementation science, governance, domain knowledge, etc.
- Reflective, predictive, adaptive, including norms, policies, goals (double-loop)
- Examples in wild: emerging LHSs, advertising, social credit, disinformation

Participant-oriented view (Shneiderman)



Feedback learning loop view



Rationale for & Scope of CSLS Research Program

- Today, practice outstrips science, engineering, and design knowledge
- Beyond descriptive/predictive theory to prescriptive/interventionist design
- Principles, methods, technologies for CSLS understanding and development
- With impact seen in major improvements in many critical societal systems

Main questions for this workshop

- **Two main questions for this workshop**
  - What are the critical gaps in knowledge: in theory, systems/technology, applications?
  - What kinds of approaches might power & characterize CSLS field? (What unique points of research & technical leverage could enable a transition to a discipline of CSLS design?)
- **Two additional questions to consider**
  - What methods should be used to evaluate impact of systematic approach to CSLS design against baseline of in-the-wild, ad hoc CSLS elements that we see already emerging today?
  - What are key features in a research roadmap to address uncertainties and to get to scale?

Overview of ideas submitted ahead of time

- Solicited inputs from all participants
- Enumeration of main topics
- Details of individual contributions
- To help bootstrap discussion, not a final structure

Enumeration of main topics

- CSS and to some extent CSLS are already happening and successful
- Motivational structures for technology-mediated social participation in CSLS
- Cyber-social mechanisms for large-scale coordination of complex tasks
- Quality and analysis of data acquired from online social networks/sources
- Humans as sensors and (noisy/biased) information channels/transducers
- Social media exploitation tools as CSLS (distillation, rumor/bias countering, cultural understanding, voice to disenfranchised, cultural understanding)
- Greatly elaborated theories of learning to underpin theory of learning in CSLS
- Principles for integration of human and AI/robotic elements into CSLS

Enumeration of main topics

- Organization of computable knowledge at scale for decision support at scale
- Closed loop and person in/on loop system design and evaluation
- Ethical, legal, regulatory aspects of statistical/heuristical deciding machines
- Design for & measurement, control, & assurance of critical system properties
- Designing CSS for ease of observation, experimentation, and adaptation
- Foundations for multi-disciplinary theory- and data-based system modeling
- Dynamics of public opinion as enabler of action based on knowledge

Preece

- Motivation for long-term participation (face to face and online) in volunteer organizations (citizen science) [Preece]
- How to engage people and ensure quality of data acquired through online crowdsourcing
- Participatory design of (and related approaches to) CSLS design, particular as these systems incorporate ML, topic modeling, sentiment analysis, etc

Griswold

- Human-robotic interaction as grand challenge
  - CSLS are already evolving in the wild
  - How are people going to be parts of these systems?
  - Human idiosyncrasies but also unique strengths => must be involved
  - Human-in-loop (often error prone) vs computer-in-loop (better, e.g., lane alerts), but neither model is ideal.
- Data representation, interpretation, and integration
  - Within but especially across domains, e.g., water/food/energy, or health(care)/education
  - E.g., databases on data.gov come in huge diversity of forms and often without sufficient meta-data to support high-confidence interpretation of meaning, reconciliation, integration
  - Progress needed for radical transparency audacious goal for smart/connected communities

Silverstein

- Knowledge representation and construction
  - Methods for assembling computable, self-explaining knowledge at scale, and doing so to enable decision-support to individuals at scale
  - Methods for assembling sensitive data into usable forms that maintain source verification sufficient for purpose
- Problem-driven research in the healthcare domain
  - Testbeds that link EMR with other data to support decisions by patients and physicians
  - Patient-similarity-based systems to support decision-making at point of care
  - Data collection systems that integrate wide sets of inputs into clinical data sources
  - Closed loop and person in/on loop system design and evaluation for different environments (inpatient, outpatient, surgical, home care, chronic monitoring, alerting, etc)
- Societal acceptability of imperfect learning, inference, and actions
  - Ethical, legal, regulatory aspects of machines that make decisions statistically/heuristically

Shneiderman

- Background
  - Great success stories of very large scale collaboration
  - Reader-to-leader theory of motivational structures for technology-mediated social participation
- Concepts, methods, tools for building CSLS for large-scale collaboration
  - Deliberative system design (ideas, plans, consensus, resources, commitments)
  - Large-scale teamwork monitoring systems (1000+, tasks, progress, evaluation/mentoring, problem identification, repair, evolution of guidelines)
  - Collaboration toolkits (checklists, communication tools, shared artifacts, tasks/schedules, remote work support)
  - Online dispute/conflict resolution (e.g., PayPal, online mediation, modern Delphi methods)

Abdelzaher

- Background
  - Social network data as a new cyber-social lens on world
  - Reliable exploitation of such noisy data to improve understanding of complex world states often involving multi-national or foreign cultures
- Gaps: Lack of holistic, interdisciplinary approach to modeling humans as sensors and social media exploitation tools as CSLSs
  - Incorporate language ambiguity into social channel bit error rate models
  - Is current information theory adequate or do we need new theory
  - Summarization as lossy compression; how does it affect error rate

Abdelzaher

- Approach: Information & estimation of human/social sensors/channels
  - human-perturbed channels => correlated errors
  - Information carrying capacity
  - Bounds on error variance of channel estimators
  - Quantification of confidence in inferences
  - Bounds on accuracy of classification, regression, reasoning, etc
- Impact
  - Media analysis systems for distillation, rumor and bias countering, cultural understanding, giving voice to disenfranchised, physical and cultural situation understanding

Fischer

- Clarify relationships with related programs
  - E.g., Cyberlearning and Future Learning Technologies
- Greatly elaborated analysis of learning to underpin theory of CSS learning
  - Much more than education
  - Similarities and differences between human and machine learning
  - Detailed analysis of machine learning, e.g., adaptive (autonomous) vs. adaptable (guided)
  - Design methods for learning systems, e.g., meta-design, libertarian paternalism, end-user dev
  - Compilation of small collection of articles, books, prototypes to help build common vision
  - Models for academic-industry collaboration around systemic problems of future

Marchionini

- Start with theories of human learning and posit whether and how these distinctions operate in a health care system, smart community, school, etc.
  - There must be differences in how we can instrument, guide, and assess learnings
  - There is a wide chasm between understanding micro processes like oxygen flow or neural firings and making a good life decision. What are the analogs for systems that learn? Surely there are micro process things to study as well as macro effects.
  - Consider memory. Retention in wetware vs retention in the cloud must have some unique metadata requirements.
  - Consider metacognition. When an individual fails to learn something or learns it poorly, the consequences may be severe. When a system fails to learn something or learns it poorly, there are also consequences. Failing to consider side effects in social media is like failing to consider side effects in clinical trials. The clinical trial phases regimen in use today is an example of a learning system that demands enormous resources. How is this like or different from a machine learning algorithm that uses reinforcement parameters to adapt as new data arrive?

Marchionini

- What are the monitoring processes, logs, check lists, and systematic assessments used in ML/NN-driven processes? How do we evaluate their use? How do the micro process monitors influence the overall system performance?
- At more general system levels, how do we assess the summative and integrative effects of hundreds of these individual learning processes to understand the effectiveness of an entire smart community/health care system?

Van Houweling

- We have to develop a deeper understanding of the social and political factors that can impact adoption of knowledge developed through learning systems
- We now live in a world where every fact has an "anti-fact," and where every expert opinion is opposed by another "expert"
- How can the knowledge obtained through learning systems be deployed in arenas where ideological and political controversy is the norm and progress requires broad consensus?
- We are much closer to developing the technical tools than those to energize the social sides of the systems we hope to improve.
- We need research which allows us to understand dynamics of public opinion formation in today's environment and its impact on individual belief and action

Peurach

- Research to understand potential for CSLS principles, methods, and tools to add value to the *improvement science* movement in education, particularly in relation to standing grand challenge problems in education
    - A main movement for last 25 years has been the *accountability* movement: of schools for student outcomes, districts for school outcomes, and innovators for effectiveness outcomes
    - A second movement (5-10 years) focuses on principles and practices of continuous learning & improvement based on new ways of working and collaboration, a.k.a., *improvement science*
    - Develop continuous L&I within schools, then network them into “learning systems”
- Note current swing in Federal priorities from outcomes to consumer choice
- [Same idea could be applied to *implementation science* in healthcare]

Other

- Design for, & measurement, control, & assurance of critical system properties
  - Integration of traditional assurance, AI/learning, and complex, large-scale social phenomena
  - New property definitions and tools/methods for monitoring, testing, prediction, assurance
  - Fundamental socio-technical system architectures and infrastructure for property assurance
  - Integration of cyber-social learning, control, evolution across spatial and temporal scales
- Designing CSS for ease of observation, experimentation, and adaptation
  - A/B testing easy in information systems; how to do it well in human-intensive systems?
  - Modular socio-technical systems architectures, layering, coupling and cohesion, options, etc
  - Assessment of complexity, technical debt, in structure and dynamics of cyber-social systems
  - What are limits on adaptive capacity of CSLS; optimal strategies for investing in learning?
- Foundations for acquiring, relating, using models from multiple disciplines
  - System models are at heart of any feedback-driven and/or learning system
  - Learning captured to a significant degree in *evolving computable models* of what's known
  - CSLS will have to augment physics-based models with models of human/social phenomena
  - Many such models will (and will have to) be learned from data, and they will be contingent
  - How to represent, acquire, validate, integrate, and evolve CSLS system models?

Other

- Continuous accrual, aggregation and integration of relevant individual and institutional but distributed data, independent of the application domain
- Contemporaneous semantic extraction from “published" articles and reports of experiments/trials/studies along with their authentication and validation
- Continuous/contemporaneous (or “real-time”) mining of the extracted and integrated data

Main questions for this workshop

- What are the critical gaps in knowledge: theory, systems, applications?
- What fundamental approaches will power and characterize CSLS field? What points of research & technical leverage enable a transition to CSLS model?
    - E.g., Socio-technical architectures for double-loop system learning at scale
    - E.g., Externalities, measures, incentives for participation, mechanism design
    - E.g., Usability, adoptability, human/smart-tech integration, on-ramps to adopt CSLS approach?

Two additional questions to consider

- What methods for evaluating impact of systematic scientific, engineering, and design approach to CSLS against baseline of in-the-wild, ad hoc CSLS?
    - E.g., measures of quality, performance, resiliency/adaptation to changes in environment
- Key features in roadmap to address uncertainties and to get to scale?
    - Feedback for CSLS conceptualization/experimentation, e.g, fast fail, prototyping, simulation
    - Challenges of scale re: system integration, community development, complexity, legacy, etc?
    - Program concepts that provide early indicators of potential outcomes? How to assess ability to succeed in at-scale thinking of this kind?

A final issue to think about: naming of field

- “I wish you as a community could do something about the term ‘learning systems’, which is too subject to limited-, mis-, and ambiguous interpretation even by the academic research community, including NSF reviewers and managers, much less the public and press” [well informed anonymous]
- In my experience, the CSLS term really doesn’t “roll off the tongue”
- The term social might also carry some liabilities

CSLS and the Mission and Vision of the NSF

- NSF Vision: A nation that creates and exploits *new concepts in science and engineering* and provides global leadership in research and education [NSF Strategic Plan for 2014-2018]
- NSF statutory mission: To promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense; and for other purposes. [NSF Foundation Act of 1950.]
- CSLS -- a timely new concept in science, engineering, and design with potentially significant applications across many sectors of our society, the foundations of which now need to be secured through research and disseminated through education