Provably-safe interventions for Human-Cyber-Physical Systems (HCPS)

Sam Burden

Assistant Professor

Electrical Engineering

University of Washington

Seattle, WA USA

http://faculty.uw.edu/sburden

CNS #1565529

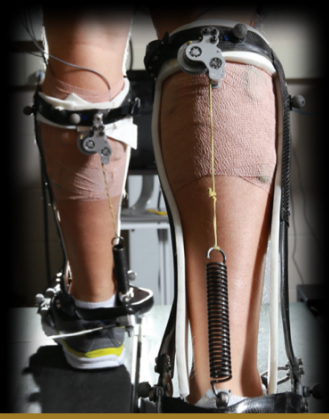
Eatai Roth

Darrin Howell

Momona Yamagami

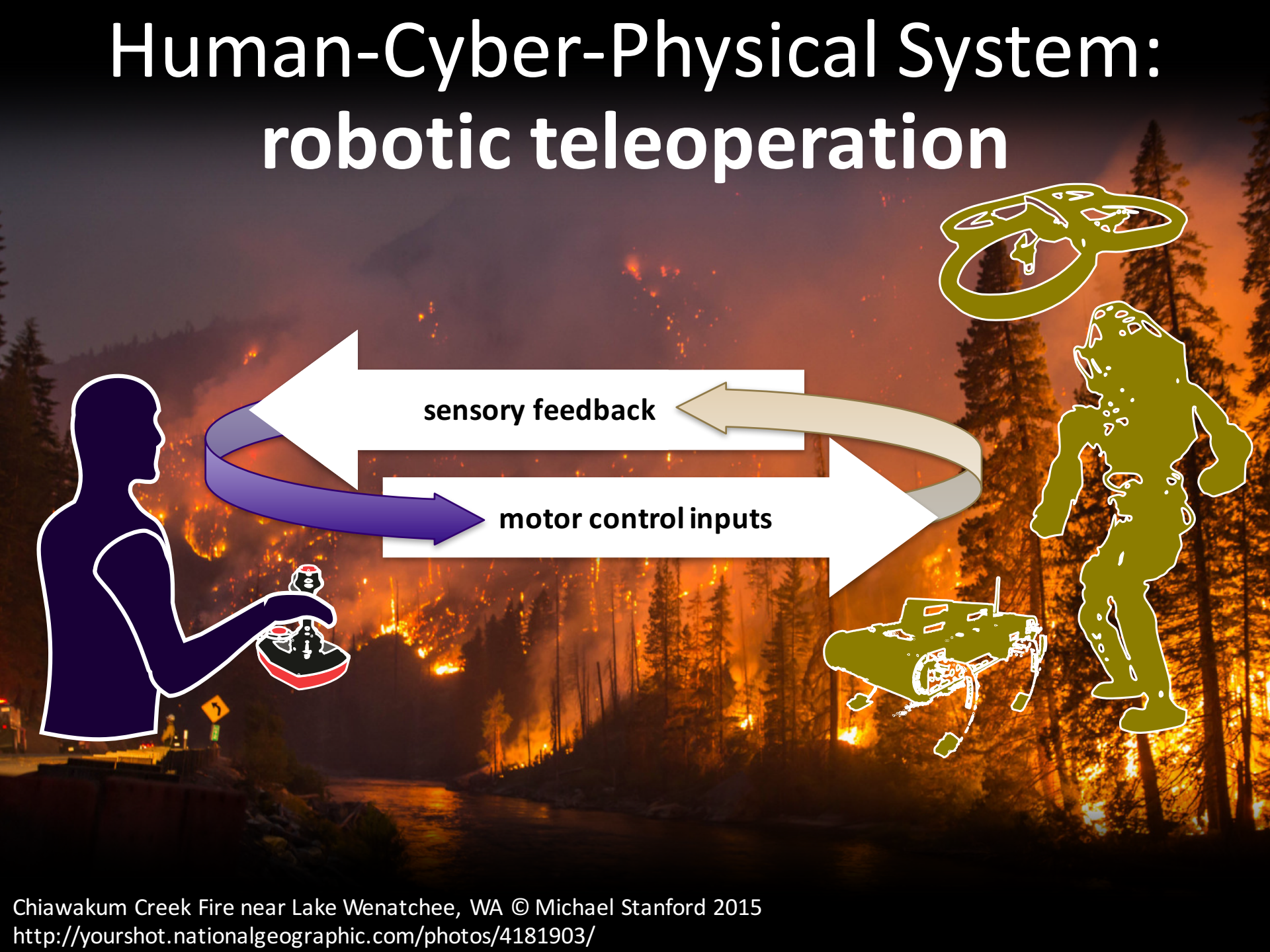
Cydney Beckwith

human interaction with the physical world is increasingly mediated by automation



Human-Cyber-Physical Systems (HCPS)

Human-Cyber-Physical System: **robotic teleoperation**

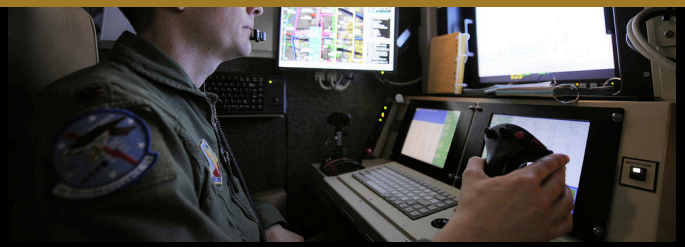
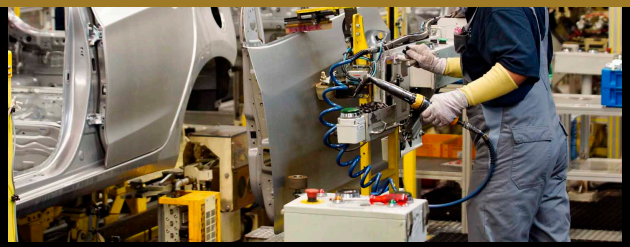


Chiawakum Creek Fire near Lake Wenatchee, WA © Michael Stanford 2015
http://yourshot.nationalgeographic.com/photos/4181903/

roles for humans and automation



legal, **ethical**, and **political** concerns
ensure humans will remain in-the-loop

Nothwang, Robinson, Burden, McCourt, Curtis *IEEE Resilience Week* 2016
The Human Should be Part of the Control Loop?

embedding humans amid automation

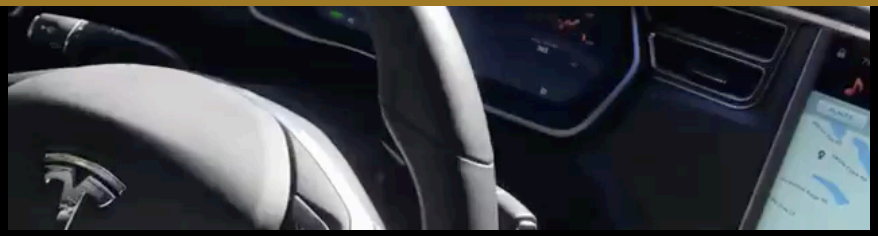
- pilot-induced oscillations in rotory & fixed-wing aircraft

  McRuer, Krendal 1974; Hess *J. Guid. Cont. Dyn.* 1997
  Pavel et al. *Prog. Aero. Sci.* 2013

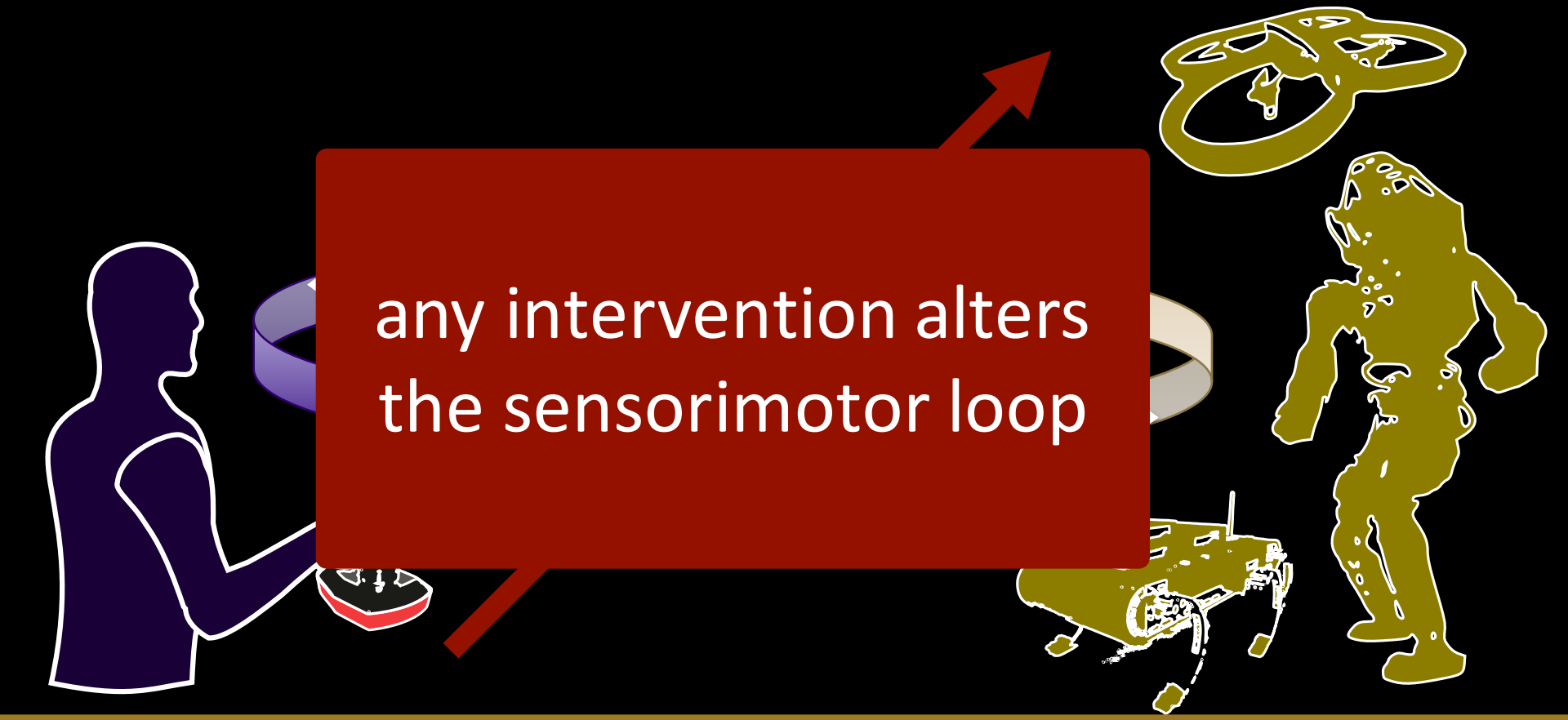
- overreliance on enhanced safety features in cars

  Rudin-Brown, Parker *Traffic Psych. Behav.* 2004;
  Moore, Zuby *Proc. Enhanced Safety Vehicles* 2013

naïve coupling can lead to performance loss ranging from mild to catastrophic



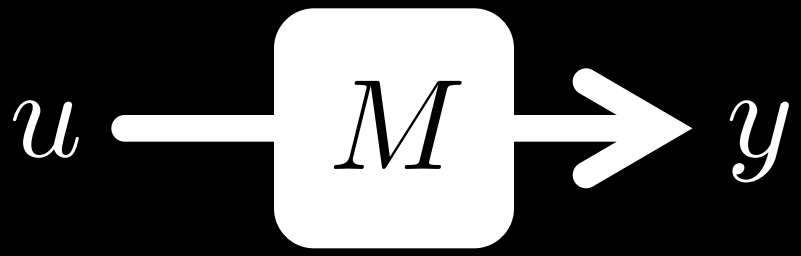
intervening in Human-Cyber-Physical Systems



safe intervention requires validated predictive models for sensorimotor loops

predictable behavior from internal models

forward model

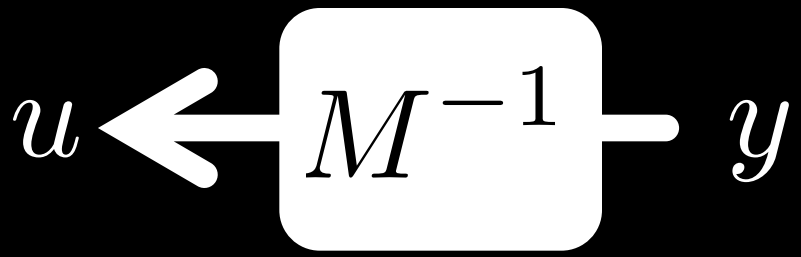


$$y = M(u)$$

- predicts sensory effects of motor actions

Sutton, Barto *Psych. Rev.* 1981
Jordan, Rumelhart *Cog. Sci.* 1992
Wolpert, Ghahramani, Jordan *Science* 1995

inverse model



$$u = M^{-1}(y)$$

- computes action expected to yield desired behavior

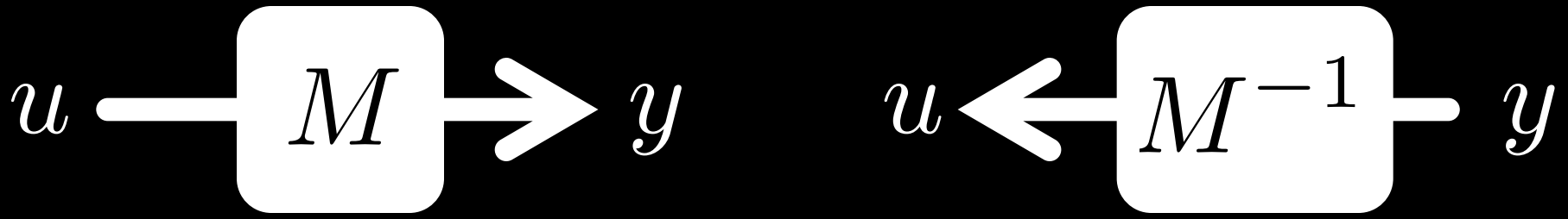
Kawato *Curr. Opin. Neurobio.* 1999
Thoroughman, Shadmehr *Nature* 2000
Conditt, Mussa-Ivaldi *PNAS* 1999

predictable behavior from internal models

- theoretical and empirical evidence for pairing of **forward + inverse** models

  Bhushan, Shadmehr *Bio. Cybern.* 1999; Sanner, Kosha *Bio. Cybern.* 1999

forward model + **inverse** model



$$y = M(u) \qquad u = M^{-1}(y)$$

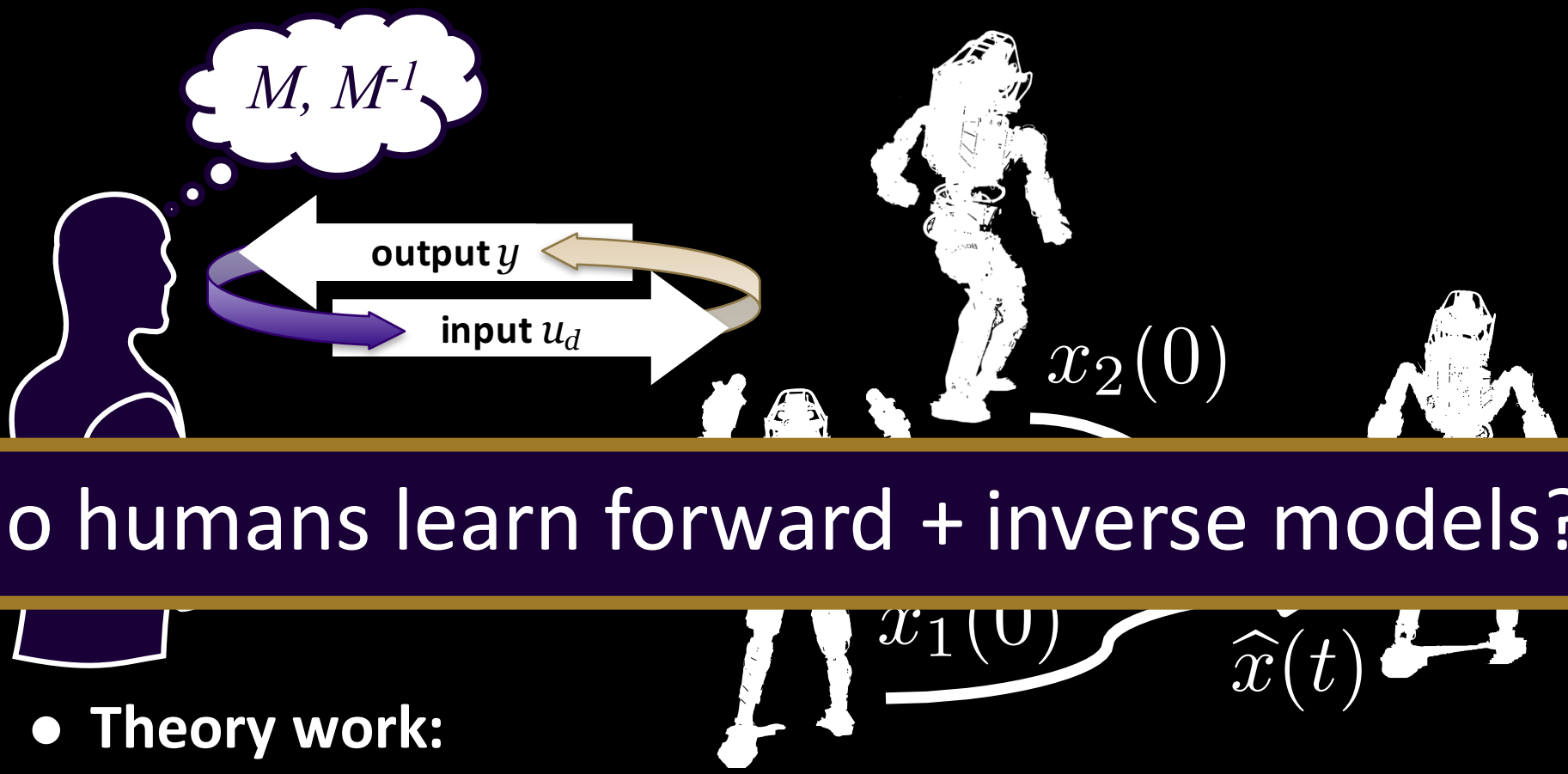
- parallels in control theory, robotics, artificial intelligence: adaptive control, internal model principle, learning

  Francis, Wonham *Automatica* 1976; Sastry, Bodson *Prentice Hall* 1989

  Sutton, Barto, Williams *IEEE CSM* 1992; Atkeson, Schaal *ICML* 1997

  Papavassiliou, Russell *IJCAI* 1999

theory for forward + inverse models

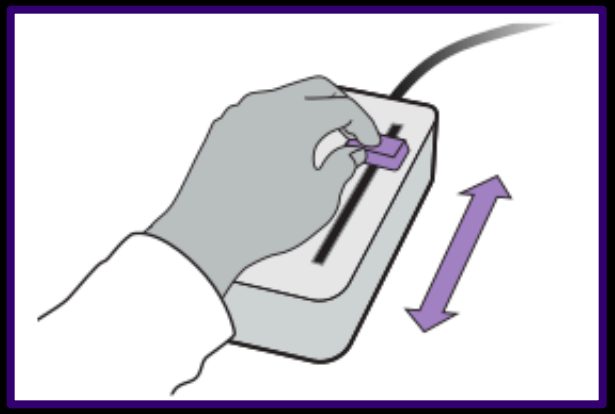


do humans learn forward + inverse models?

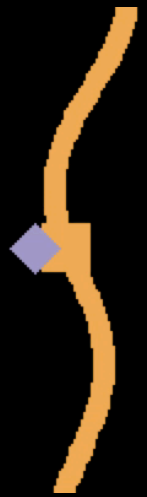
- **Theory work:**
  - for stable model pair, trajectories x_1 and x_2 converge to $\widehat{x}$
  - feedforward input "asymptotically inverts" dynamics

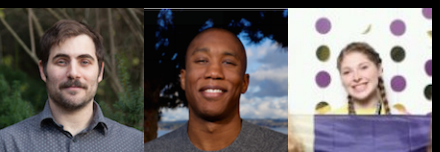
Robinson, Scobee, Burden, Sastry *SPIE-DSS* 2016
Dynamic inverse models in human-cyber-physical systems

experiments with forward + inverse models



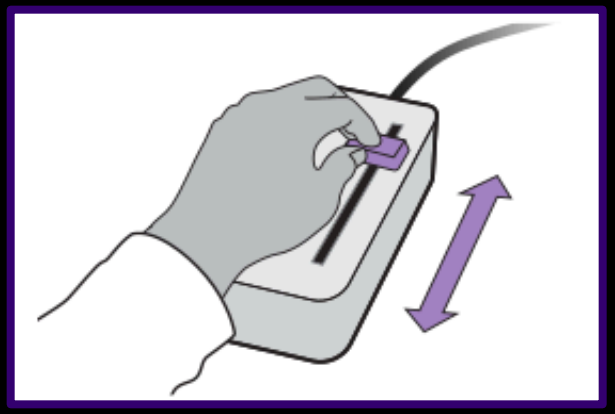
– subjects use 1-dimensional input device to control **cursor motion** to track **specified reference**



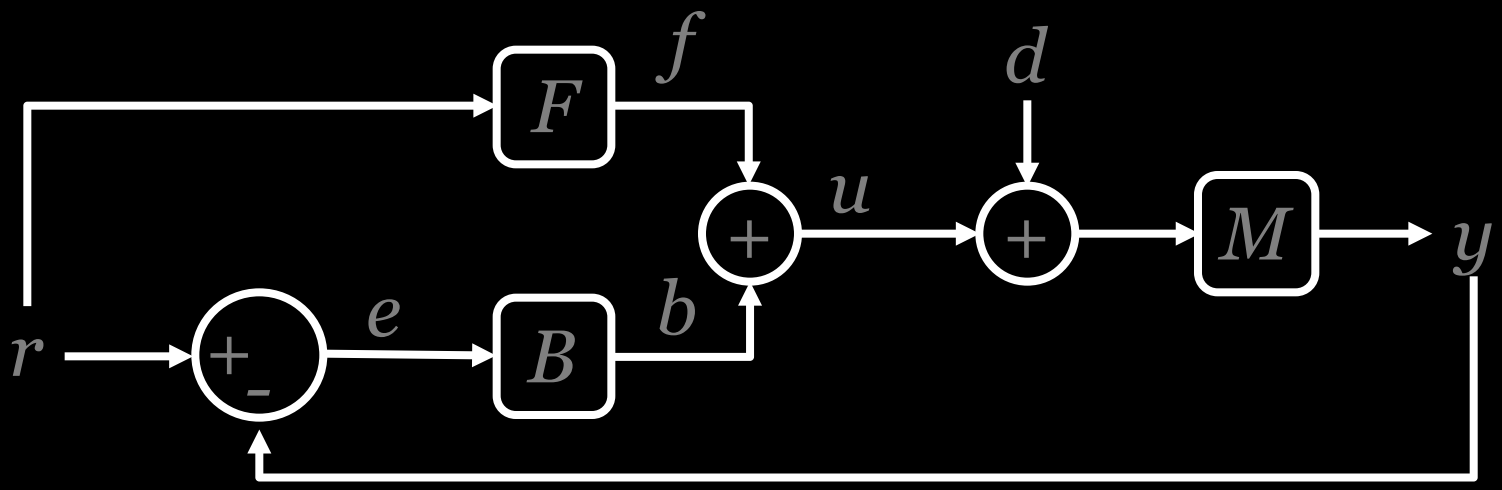


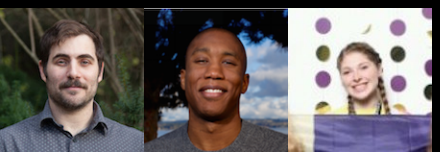
Roth, Howell, Beckwith, Burden *SPIE 2017*
Toward experimental validation of a model for human sensorimotor learning and control in teleoperation

experiments with forward + inverse models



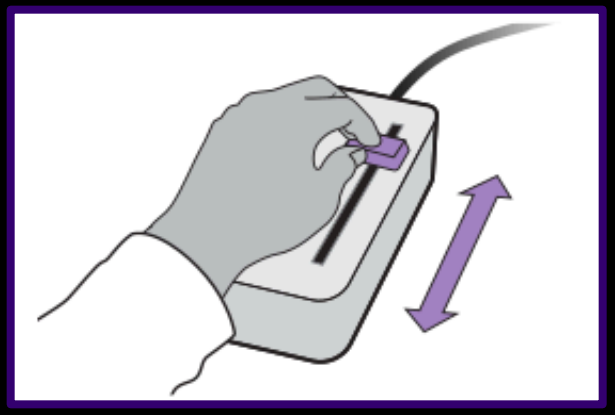
– subjects use 1-dimensional input device to control **cursor motion** to track **specified reference**



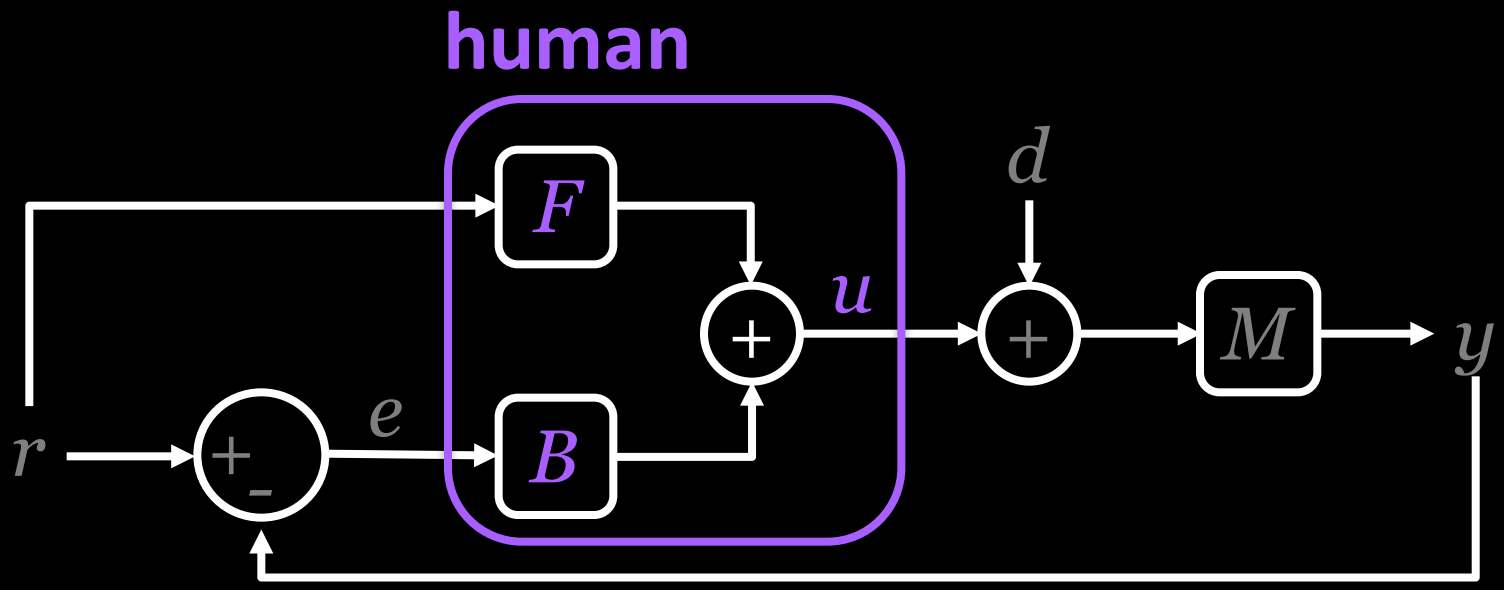


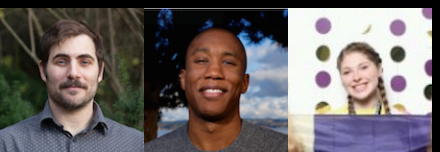
Roth, Howell, Beckwith, Burden *SPIE 2017*
Toward experimental validation of a model for human sensorimotor learning and control in teleoperation

experiments with forward + inverse models



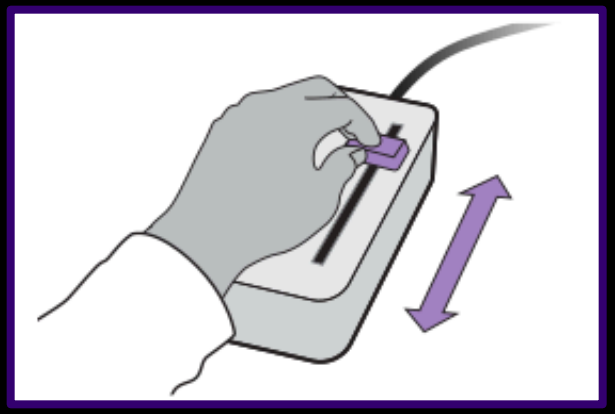
– subjects use 1-dimensional input device to control **cursor motion** to track **specified reference**



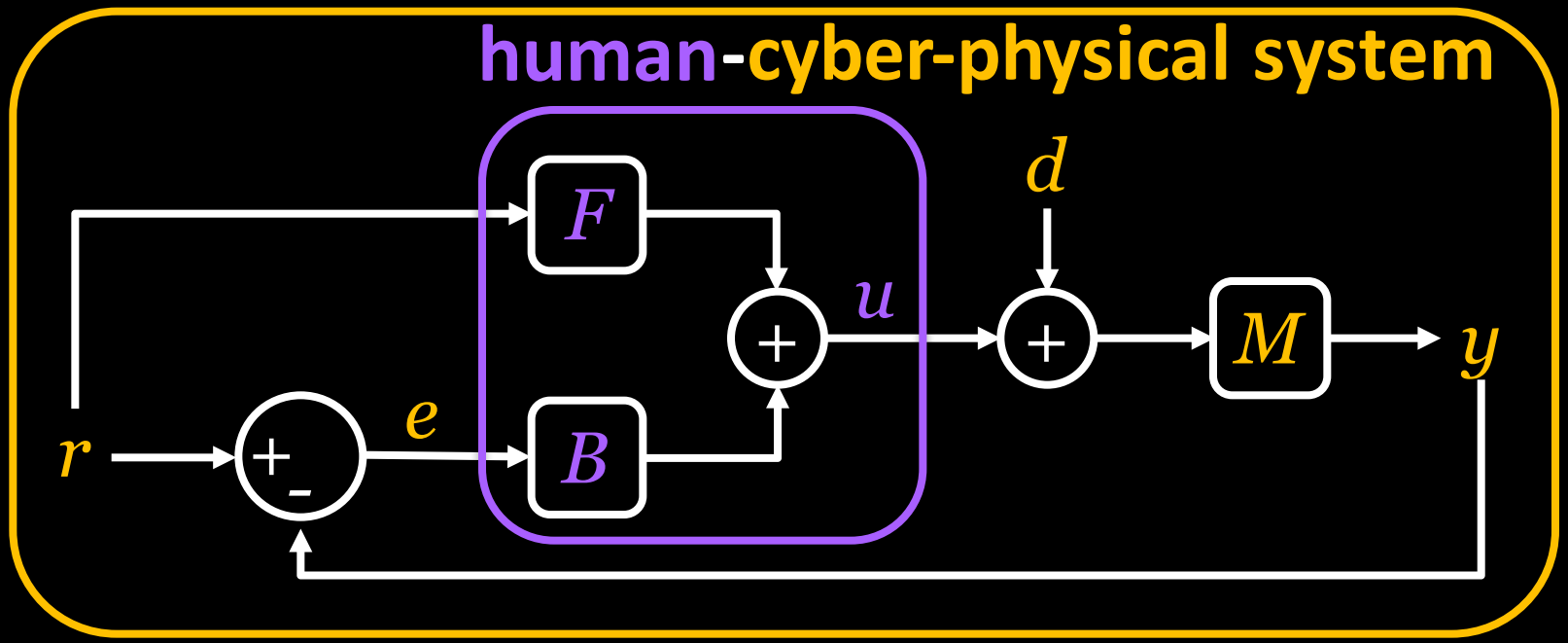


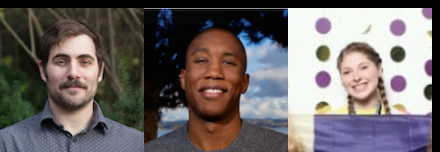
Roth, Howell, Beckwith, Burden *SPIE 2017*
Toward experimental validation of a model for human sensorimotor learning and control in teleoperation

experiments with forward + inverse models



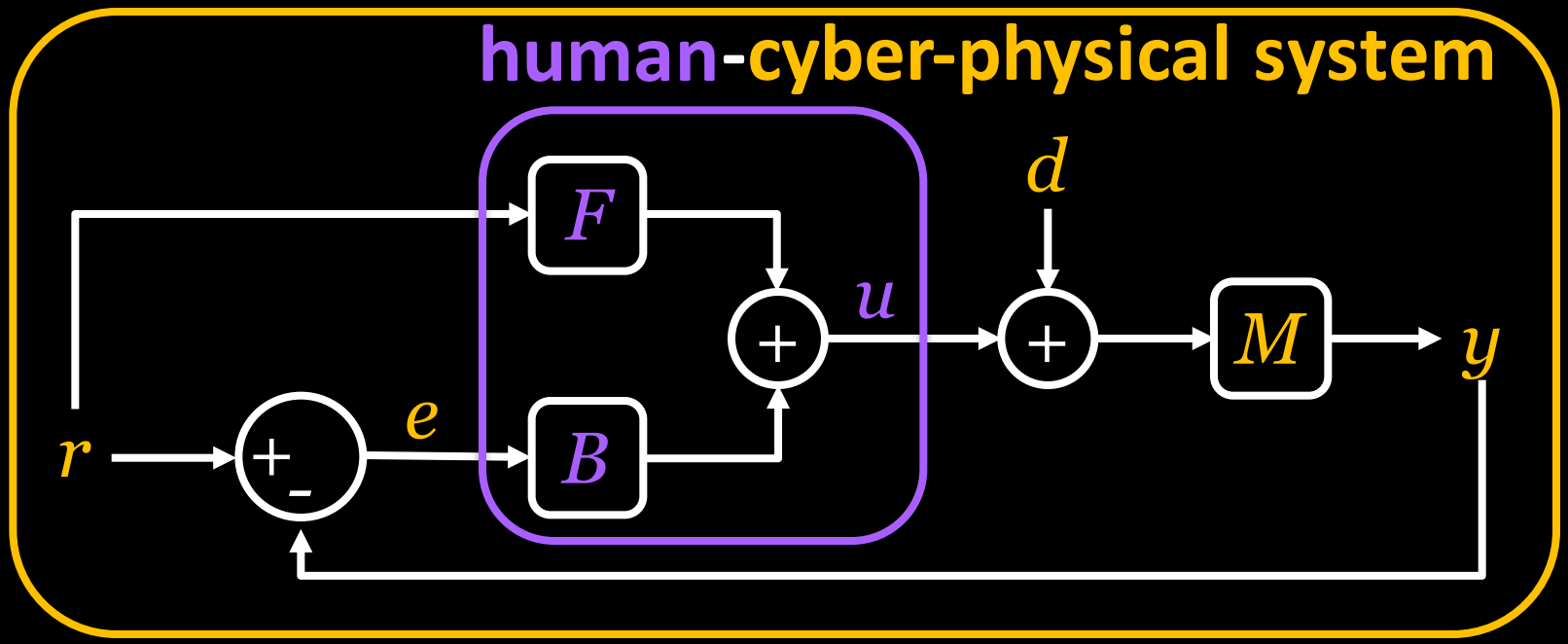
– subjects use 1-dimensional input device to control **cursor motion** to track **specified reference**



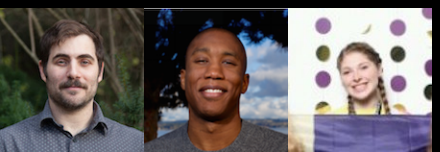


Roth, Howell, Beckwith, Burden *SPIE 2017*
Toward experimental validation of a model for human sensorimotor learning and control in teleoperation

empirically estimating learned model



- vary reference (r) and disturbance (d), estimate human feedforward (F), feedback (B)
- human learns to invert specified model (M) if feedforward approximates the inverse ($F \approx M^{-1}$)



Roth, Howell, Beckwith, Burden *SPIE 2017*
Toward experimental validation of a model for human sensorimotor learning and control in teleoperation

Provably-safe interventions for Human-Cyber-Physical Systems (HCPS)

Sam Burden

Assistant Professor

Electrical Engineering

University of Washington

Seattle, WA USA

http://faculty.uw.edu/sburden

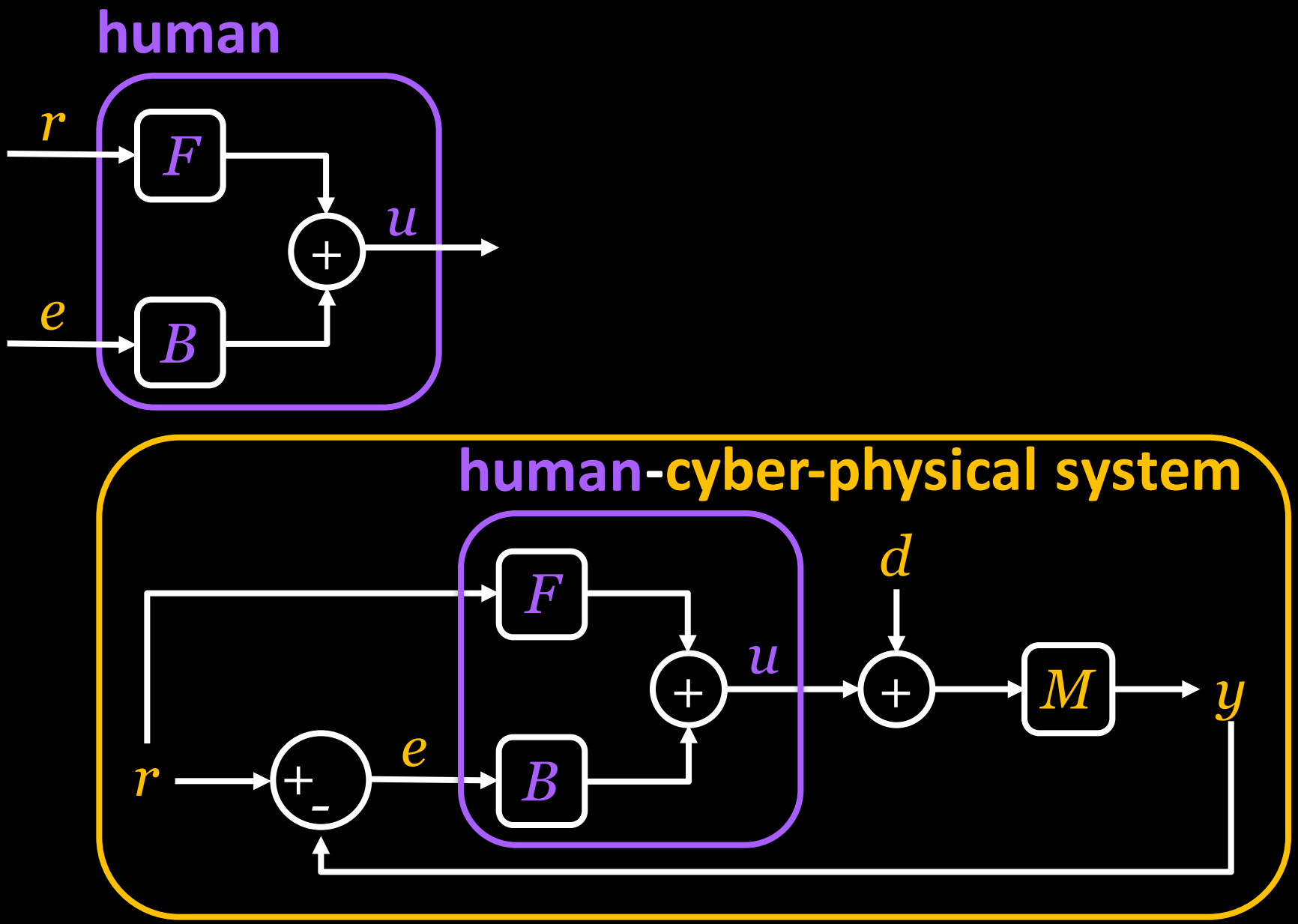
CNS #1565529

Eatai Roth

Darrin Howell

Momona Yamagami

Cydney Beckwith


human
r
F
u
e
B
human-cyber-physical system
d
F
u
M
y
r
e
B
+
-