

Mixed Initiative Interaction

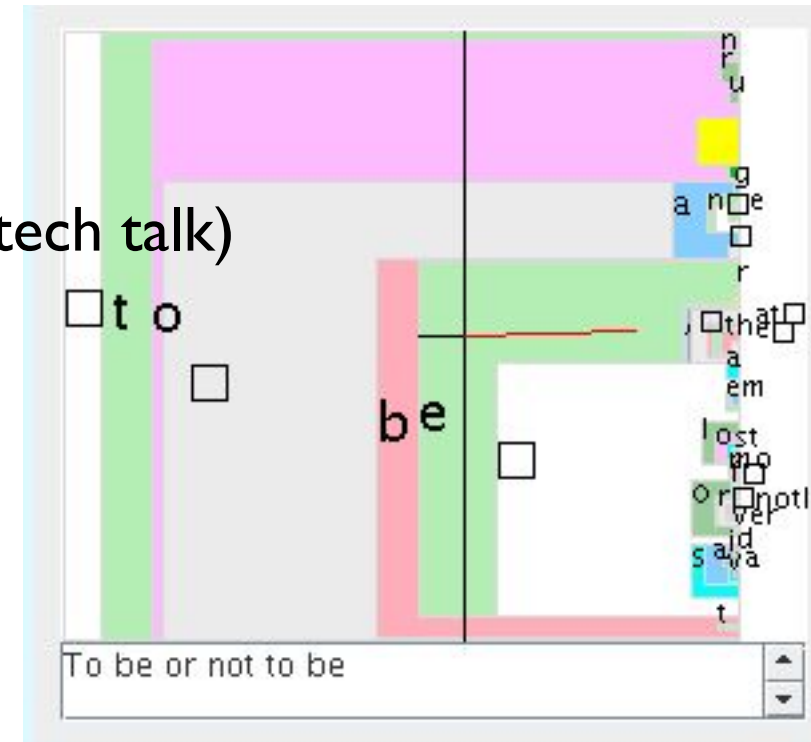
MPhil ACS module P230 - Alan Blackwell



What is Mixed Initiative?

A classic illustration of mixed initiative – predictive text

- David MacKay's Dasher was the first ML-based predictive text system
 - turned a language model visualisation into an interactive controller
 - the driving/flying metaphor can appear 'autonomous'
 - Continues generating text unless you stop it – now familiar (after 25 years!) from LLMs
 - Steers you toward most likely sequences from its training data (was this plagiarism?)
 - It took serious effort to "type" something unlikely
- Keith Vertanen's reimplementation
 - <https://dasher-site.netlify.app>
 - <https://dasher-project.github.io/dasher-web/browser/>
- David's language model explanation (extract from Google tech talk)
 - <https://youtu.be/0d6ylquOKQ0>
- Instructions on how to use Dasher by Keith Vertanen
 - <https://youtu.be/nr3s46l3DX8>
- Testimonials by users (developer Ada Majorek at 1:15)
 - <https://youtu.be/QxFEUk3J89Q>



Principles of Mixed-Initiative User Interfaces

- Classic paper by Eric Horvitz:
 - Principles of mixed-initiative user interfaces.
 - In proceedings CHI 1999, pp. 159-166.
- Advocates elegant coupling of *automated services* with *direct manipulation*
- *Autonomous actions should be taken only when an agent believes that they will have greater expected value than inaction for the user.*
 - (note who controls this relationship – “when the agent believes” – and note Horvitz’s current job as Chief Scientific Officer of Microsoft)

How to add value with automation

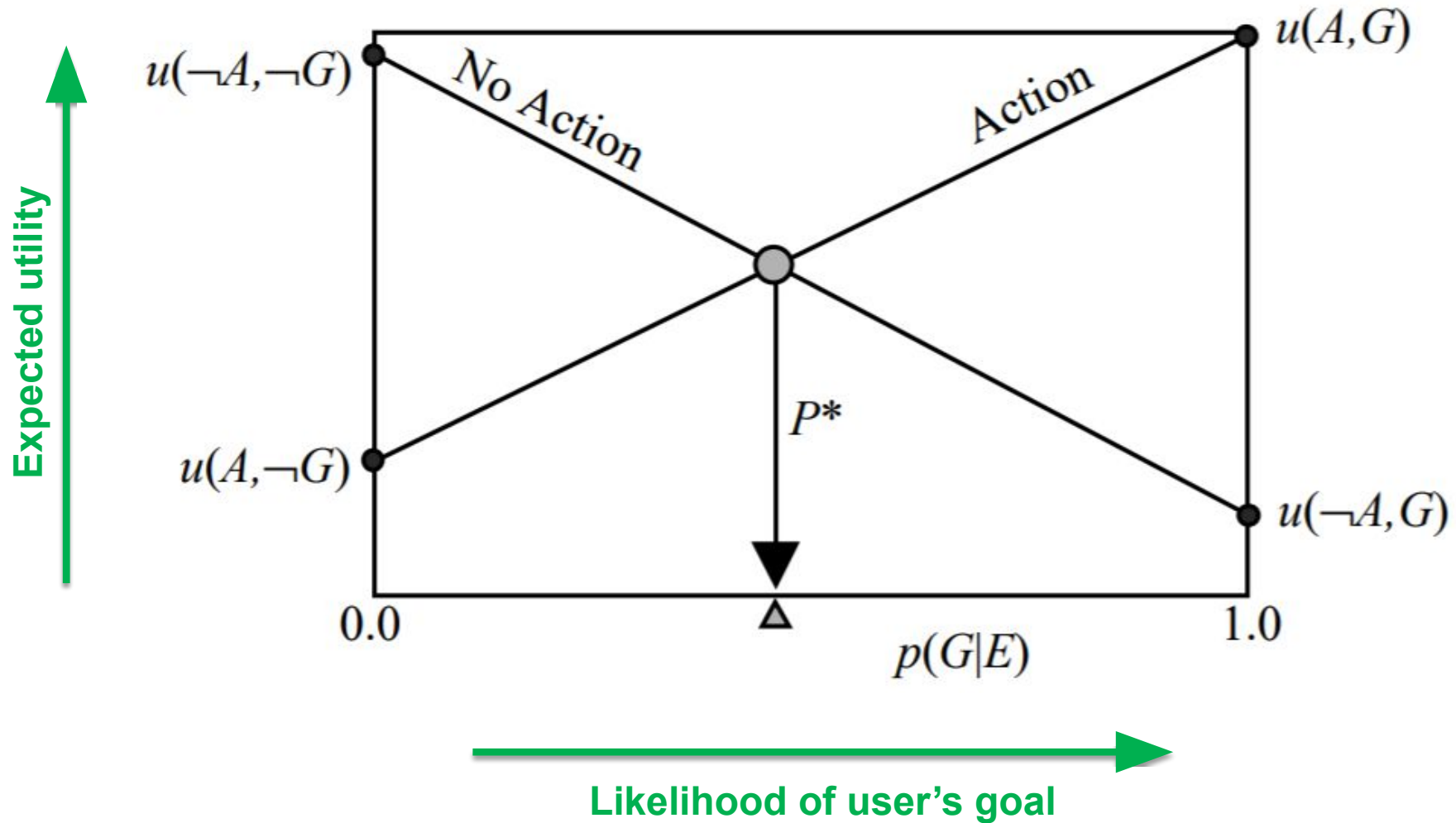
- Consider uncertainty about user's goals
- Consider status of user's attention in timing services
 - with cost/benefit of deferring action to a time when action will be less distracting.
- Infer ideal action in light of costs, benefits, and uncertainties
- Employ dialog to resolve key uncertainties
 - consider costs of bothering user needlessly
- Allow efficient direct invocation and termination
- Minimise cost of poor guesses about action and timing

Expected utility of automated action

- assume an agent can infer $p(G|E)$
 - likelihood of the user's goal being G
 - given the observed evidence E
- and decide whether to take action A , based on the utility that this action would have in achieving the desired goal G rather than other possible outcomes (not- G)

	Desired Goal	Not Desired
Action	$u(A, G)$	$u(A, \neg G)$
No Action	$u(\neg A, G)$	$u(\neg A, \neg G)$

Expected utility threshold for action vs no action



A probabilistic view of user interaction

□ Machine:

- I know how to do several things.
- I wonder which one the user wants me to do? *(a likelihood estimate)*

□ User:

- This machine can do a whole bunch of stuff.
- What is most likely to make it do the right stuff? *(a likelihood estimate)*

□ Machine:

- I think the user has *(may have)* made a mistake

□ User:

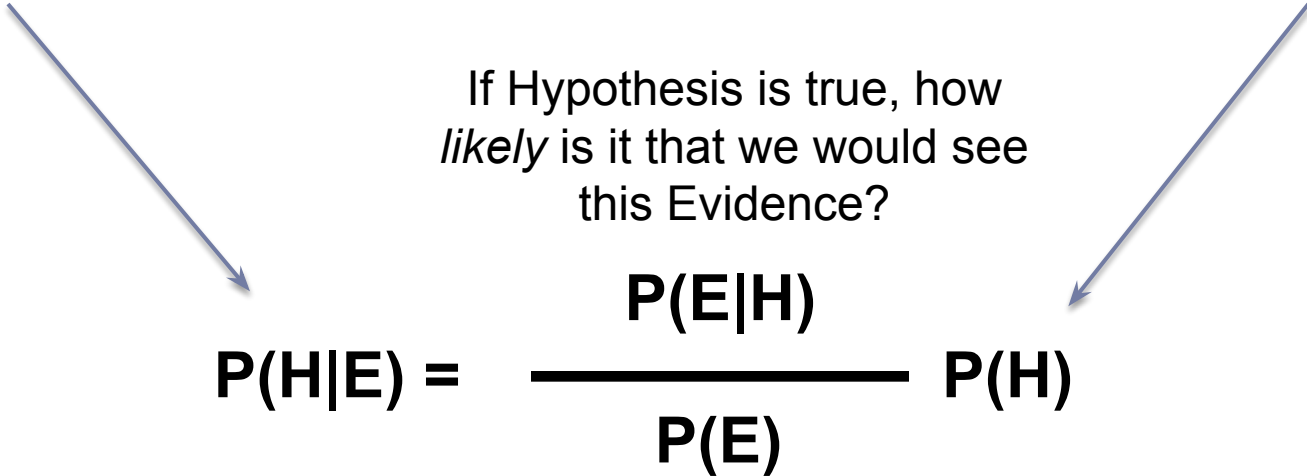
- I think the machine has *(may have)* made a mistake

Bayes theorem (for Bayesian inference)

Posterior probability of
Hypothesis *after* taking new
Evidence into account

Prior inferred probability of
this **Hypothesis** *before* new
Evidence became available.

If Hypothesis is true, how
likely is it that we would see
this Evidence?


$$P(H|E) = \frac{P(E|H)}{P(E)} P(H)$$

What is the probability of
seeing E, under all possible
hypotheses?

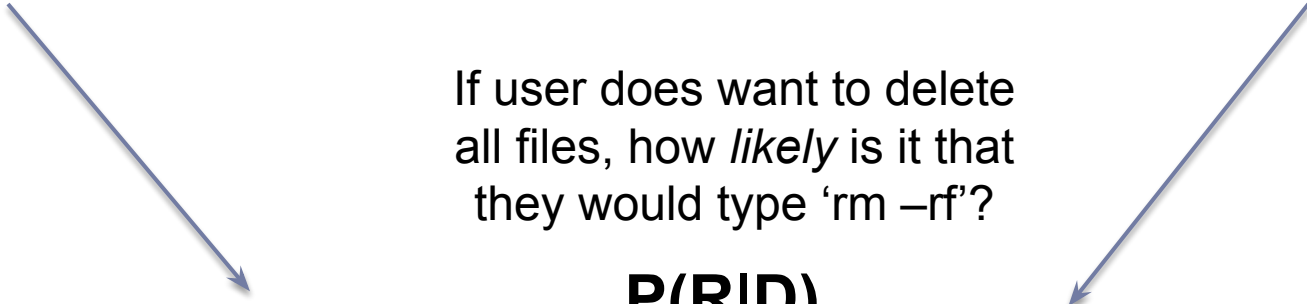
H: Hypothesis
E: Evidence

Bayesian inference inference of user intention

Probability that user wants to delete all files, given that they just typed 'rm -rf'

(Prior) probability that user wanted to delete all files *before* we saw this.

If user does want to delete all files, how *likely* is it that they would type 'rm -rf'?


$$P(D|R) = \frac{P(R|D)}{P(R)} P(D)$$

What is the probability user would type 'rm -rf', under all possible hypotheses?

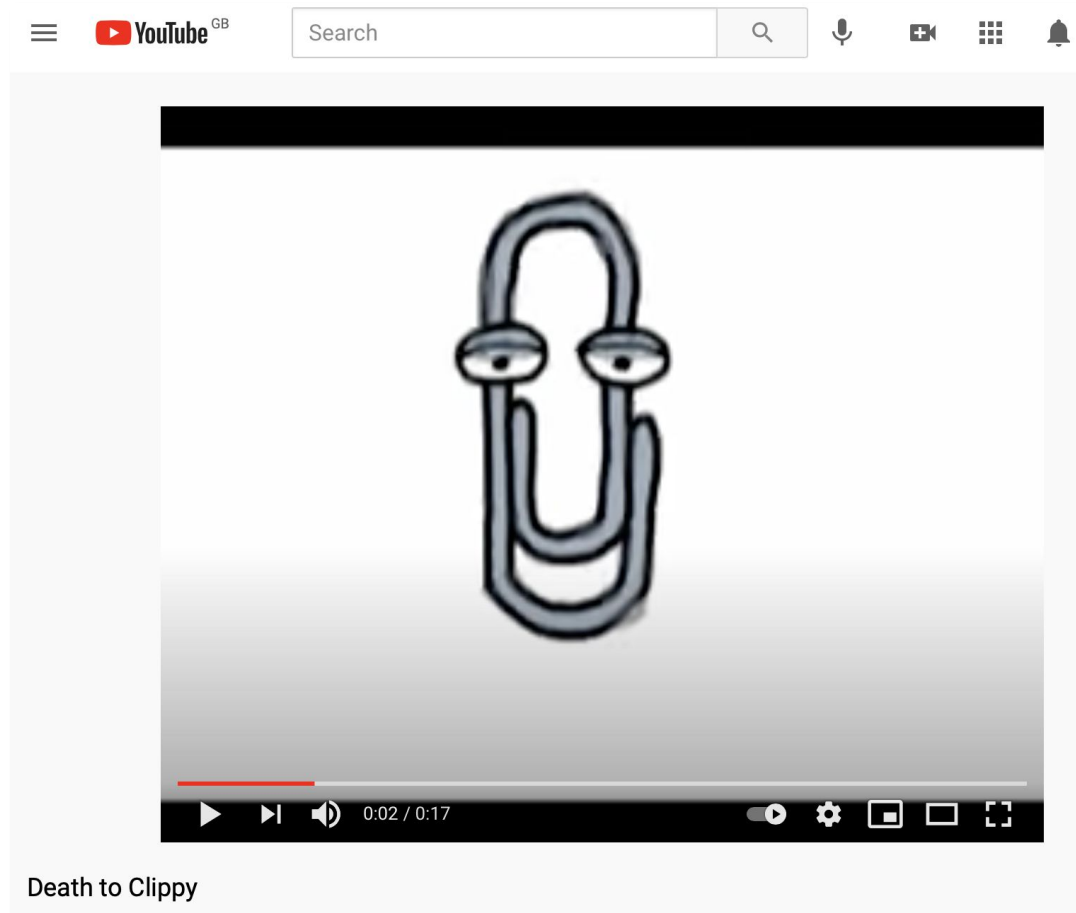
D: User wants to **D**elate all their files

R: User has typed 'rm -rf'



Another classic (notorious) example of mixed initiative

□ <https://www.youtube.com/watch?v=0ej4tW7hLkE>



It looks like you're writing a letter.

Would you like help?

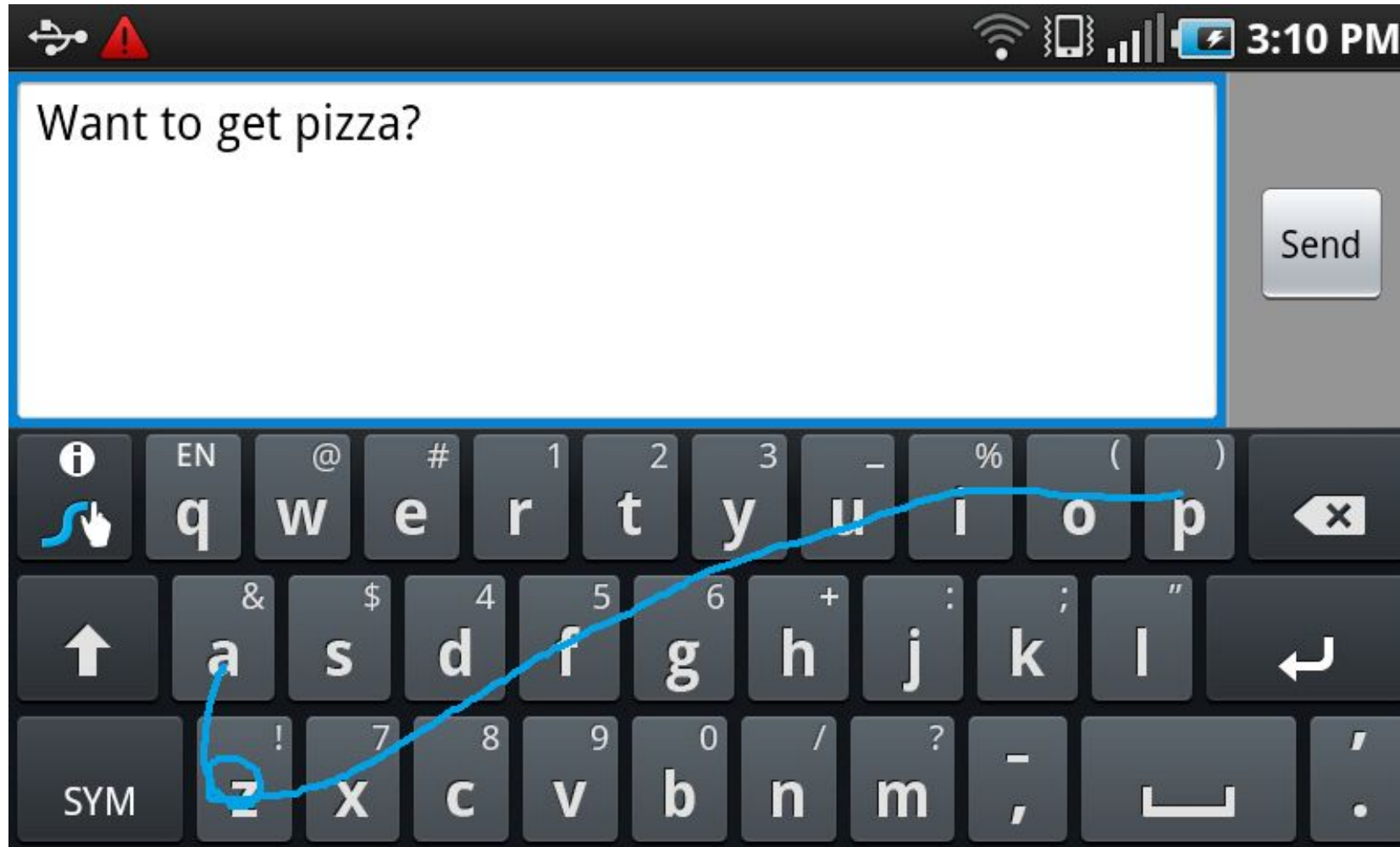
- Get help with writing the letter
- Just type the letter without help
- ☐ Don't show me this tip again



Unobtrusive direct manipulation strategy: semantic pointing



Unobtrusive direct manipulation strategy: gesture keyboard



Sometimes I just popup for no particular reason, like now.





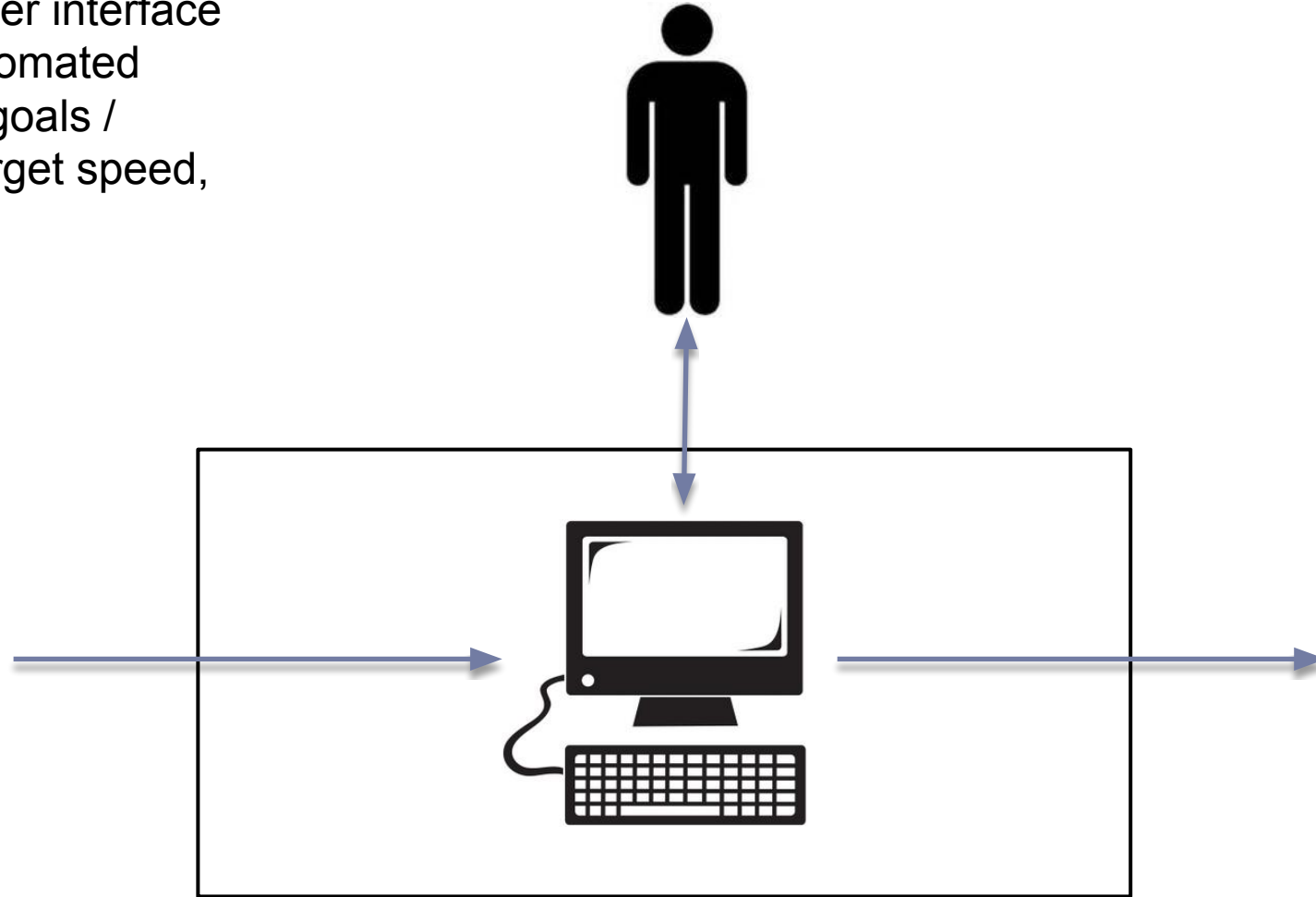
Information flow and mixed initiative

System boundaries – autonomous vehicle case

- See Shneiderman (2020) for broad background (linked to course materials page)
- Where does information enter the system?
 - User defines setpoint (“cruise control”)
 - Supplier offers features (“active braking”)
 - Regulator defines policy (“following distance”)
 - Government provides infrastructure (“lane markings”)
- Notes:
 - Even if the system includes “autonomous” closed loop control algorithms, information is acquired through more or less costly interactive processes outside the system boundary.
 - All closed loop control systems do machine learning (reacting to error signal, tuning gain and stability etc), but as interaction with such systems becomes routine, these *cybernetic* components are no longer considered intelligent.

Human in the Loop? Conventional system design

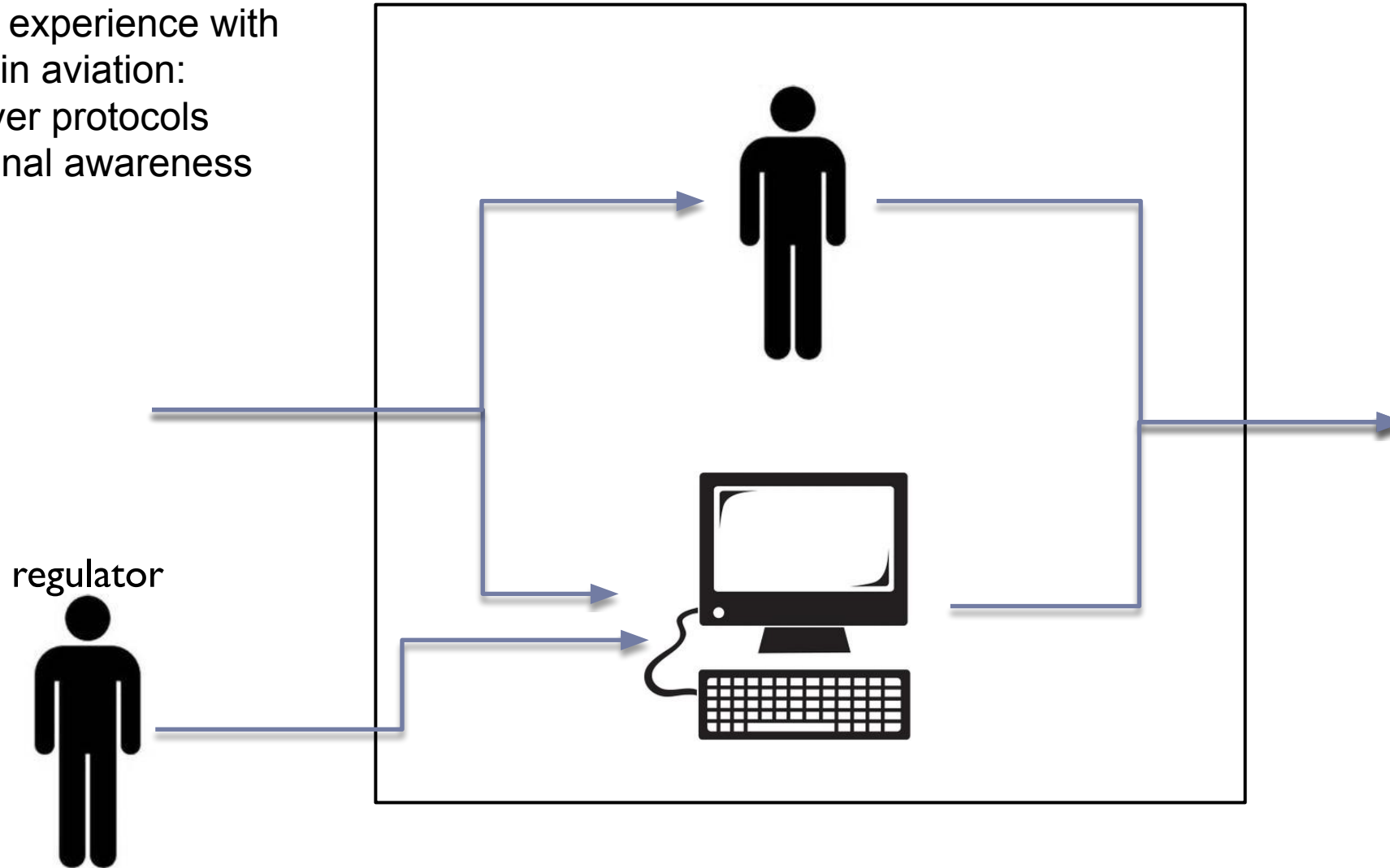
The purpose of the user interface is to configure the automated behavior (e.g. define goals / setpoints), such as target speed, emergency stop



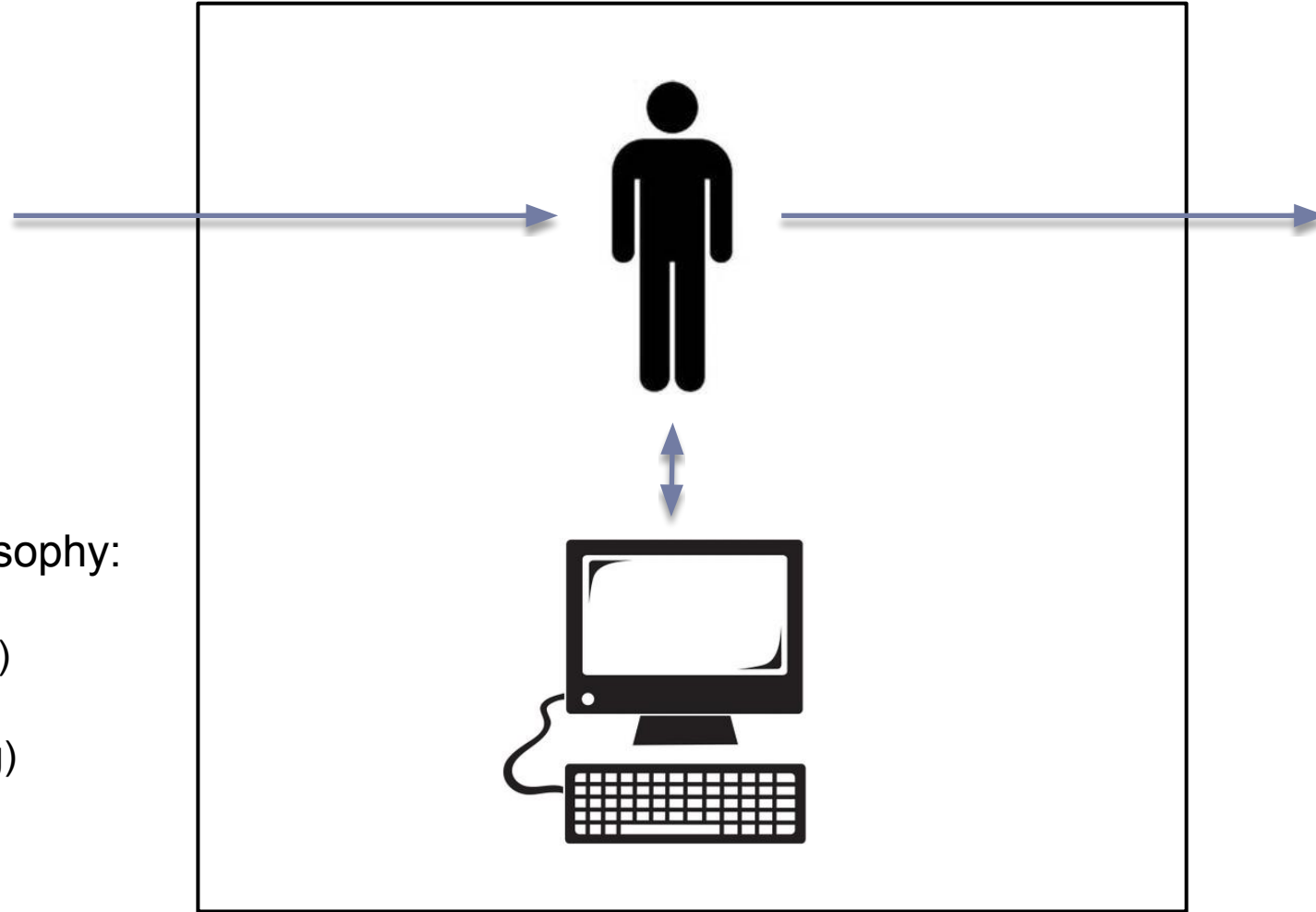
Human in the Loop? Hybrid system design

From long experience with autopilots in aviation:

- Handover protocols
- Situational awareness



Human in the Loop? Human-centric system design



The intelligent tool philosophy:

- Context-specificity
(e.g. parking assistance)
- Error avoidance
(e.g. out of lane warning)



Studying Agency and Control

- Work with David Coyle, Per Ola Kristensson and clinical collaborators

The *experience of agency* is defined as:

- *The experience of controlling one's own actions and, through this control, affecting the external world.*
- It is the experience of ourselves as agents that allows us to instinctively say:

“I did that”

Haggard, P. & Tsakiris, M., *The Experience of Agency: Feelings, Judgments, and Responsibility*.
Current Directions in Psychological Science, 2009.

Fact vs. the experience of agency

- Passivity phenomena in schizophrenia
 - People feel that their actions - and sometimes their thoughts and emotions - are not under their own control. Rather they are under the control of some external force or agent.
- Mellor reports on a patient with schizophrenia saying:

“It is my hand and arm that move, and my fingers pick up the pen, but I don’t control them.”

Mellor, C.S., First rank symptoms of schizophrenia. Br J Psychiatry, 1970.

Shneiderman's “golden rules of HCI”

Rule no. 7: “Support an internal locus of control”

This rule is based on the observation that:

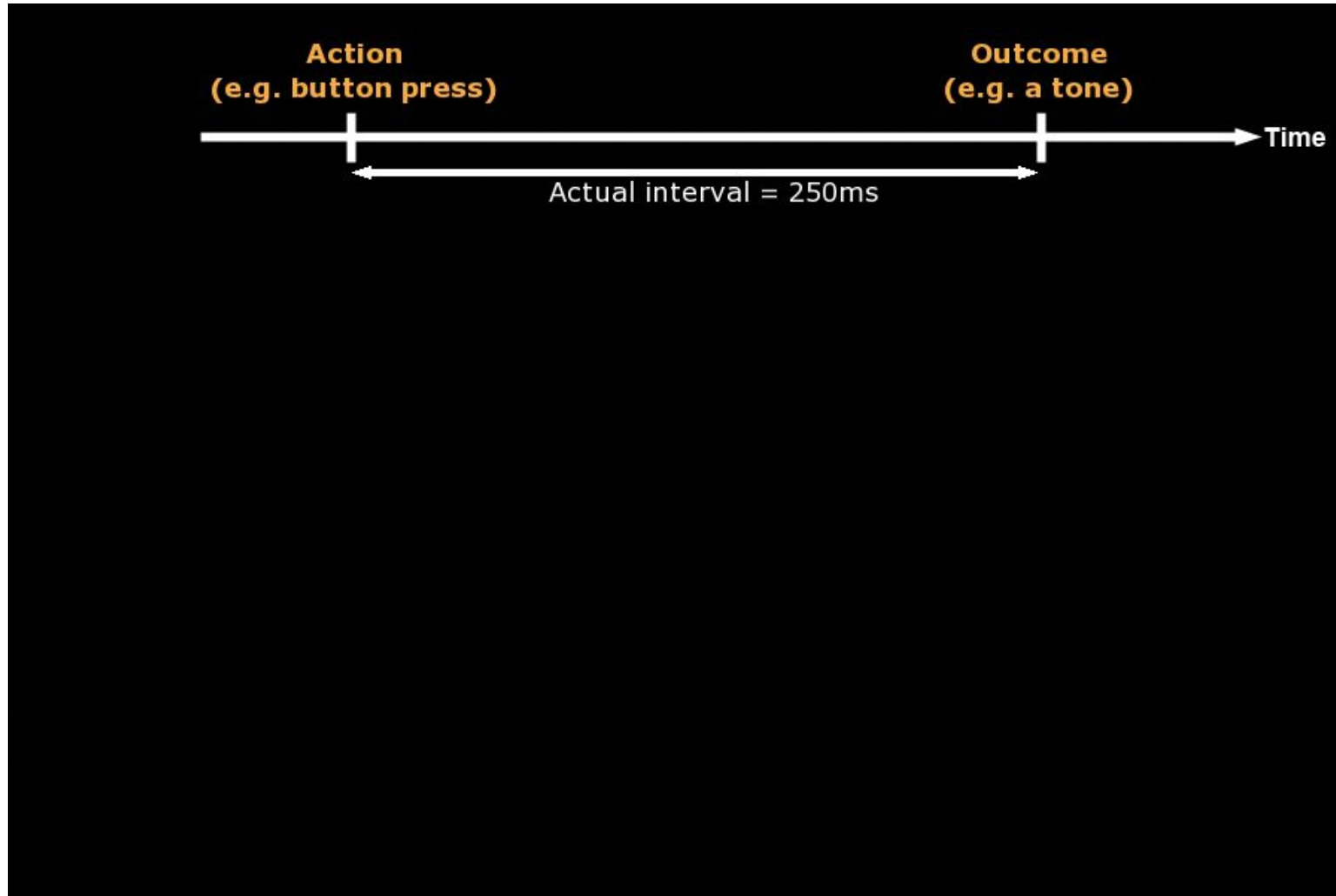
“Users strongly desire the sense that they are in charge of the system and that the system responds to their actions.”

*Shneiderman, B. & Plaisant, C. 2009
Designing the User Interface: Strategies for Effective Human-Computer Interaction.*

Developing a research agenda

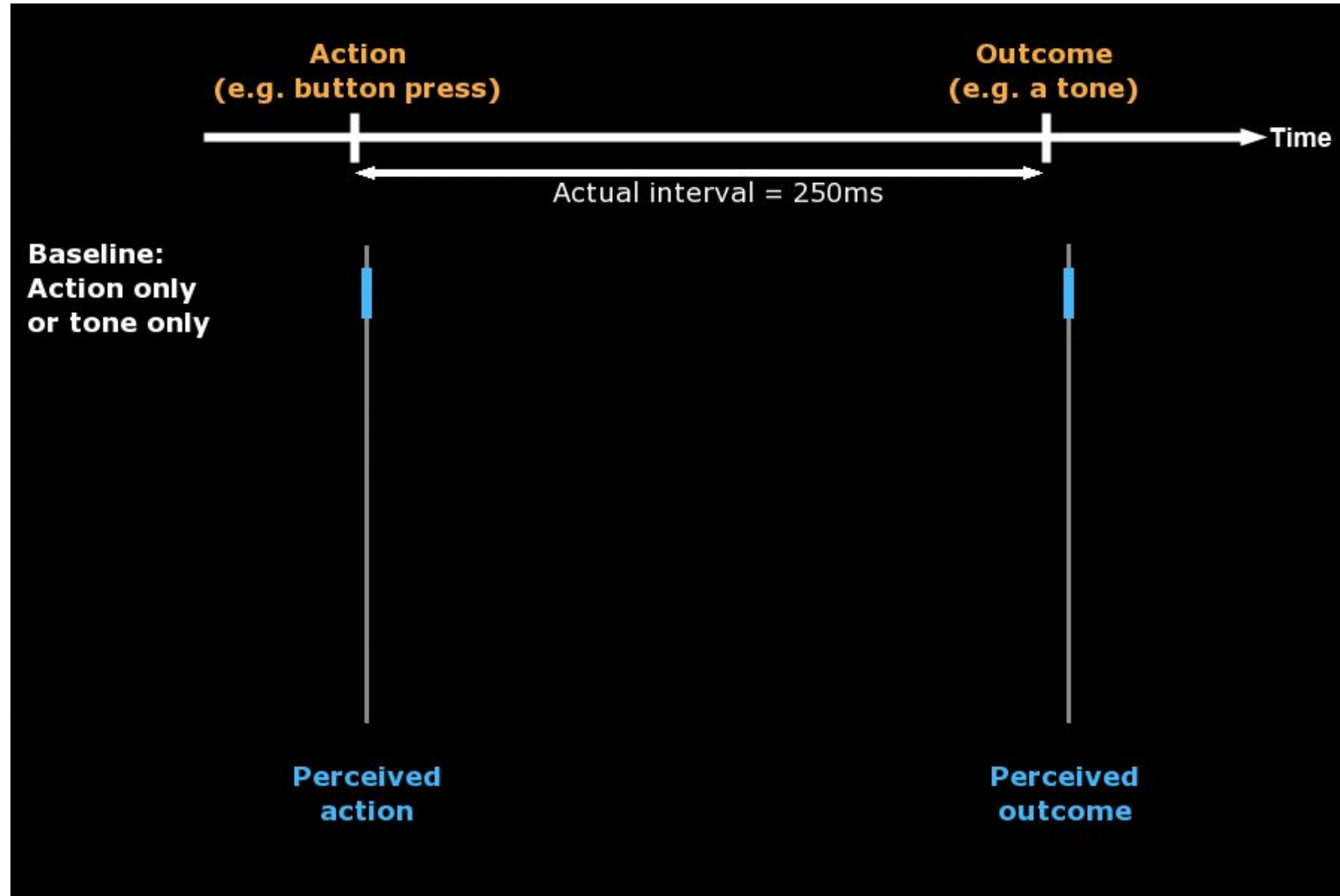
- An implicit metric to measure peoples' experience of agency.
- Two experiments that apply this metric in HCI contexts.

Intentional binding



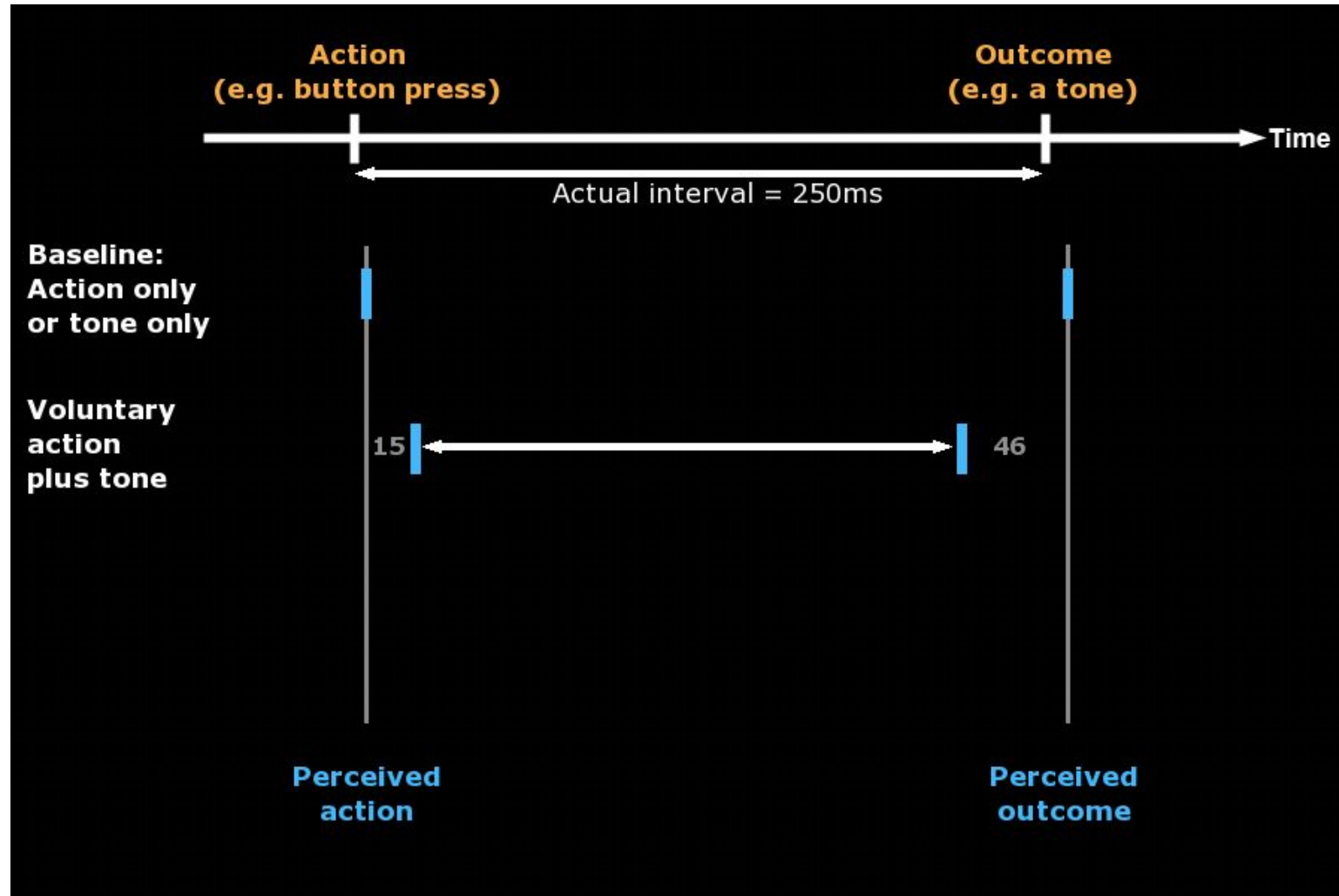
Haggard, P. & Tsakiris, M., *The Experience of Agency: Feelings, Judgments, and Responsibility*. Curr Dir Psychol Sci, 2009, 18(4) p.242-46.

Intentional binding



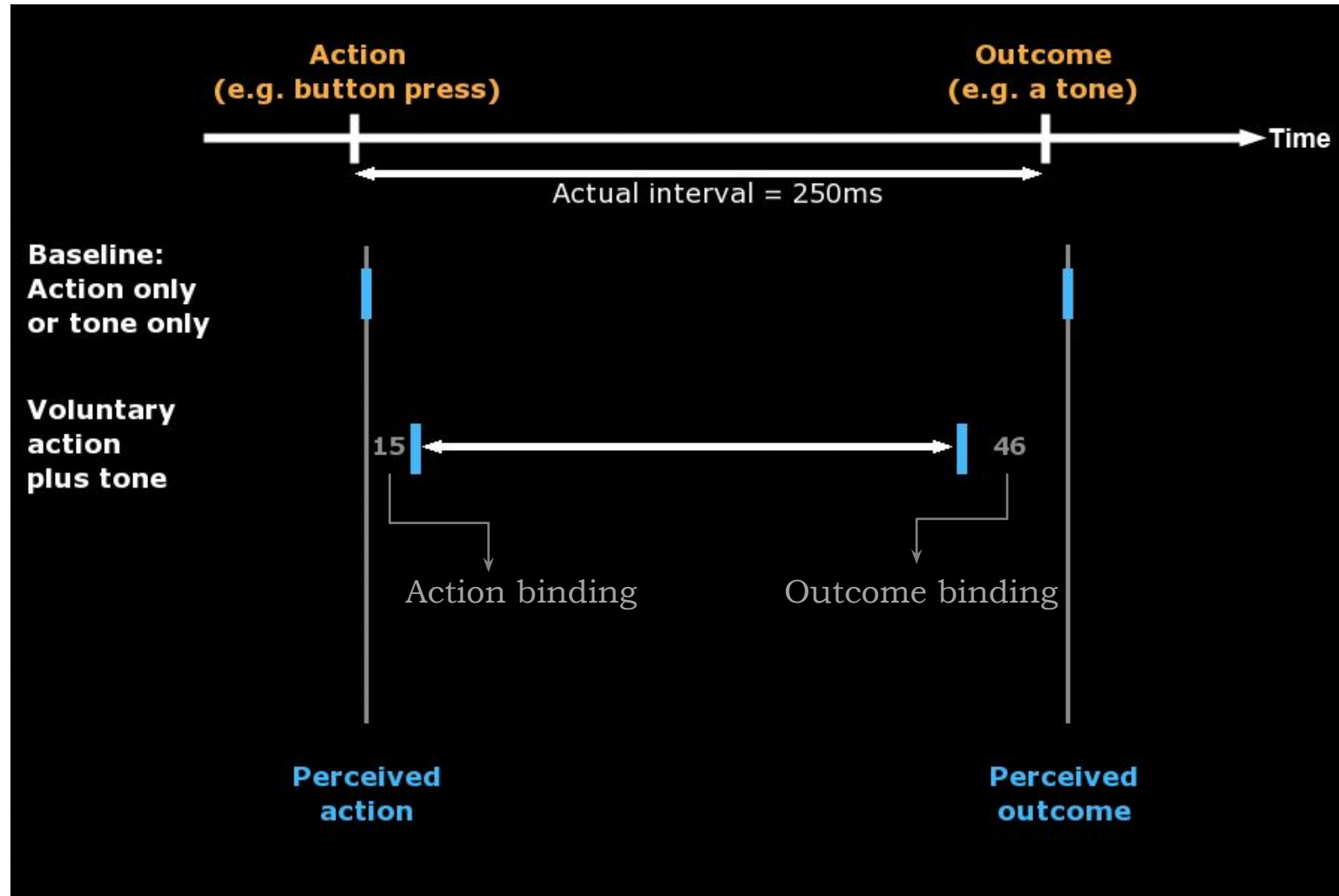
Haggard, P. & Tsakiris, M., *The Experience of Agency: Feelings, Judgments, and Responsibility*. Curr Dir Psychol Sci, 2009, 18(4) p.242-46.

Intentional binding



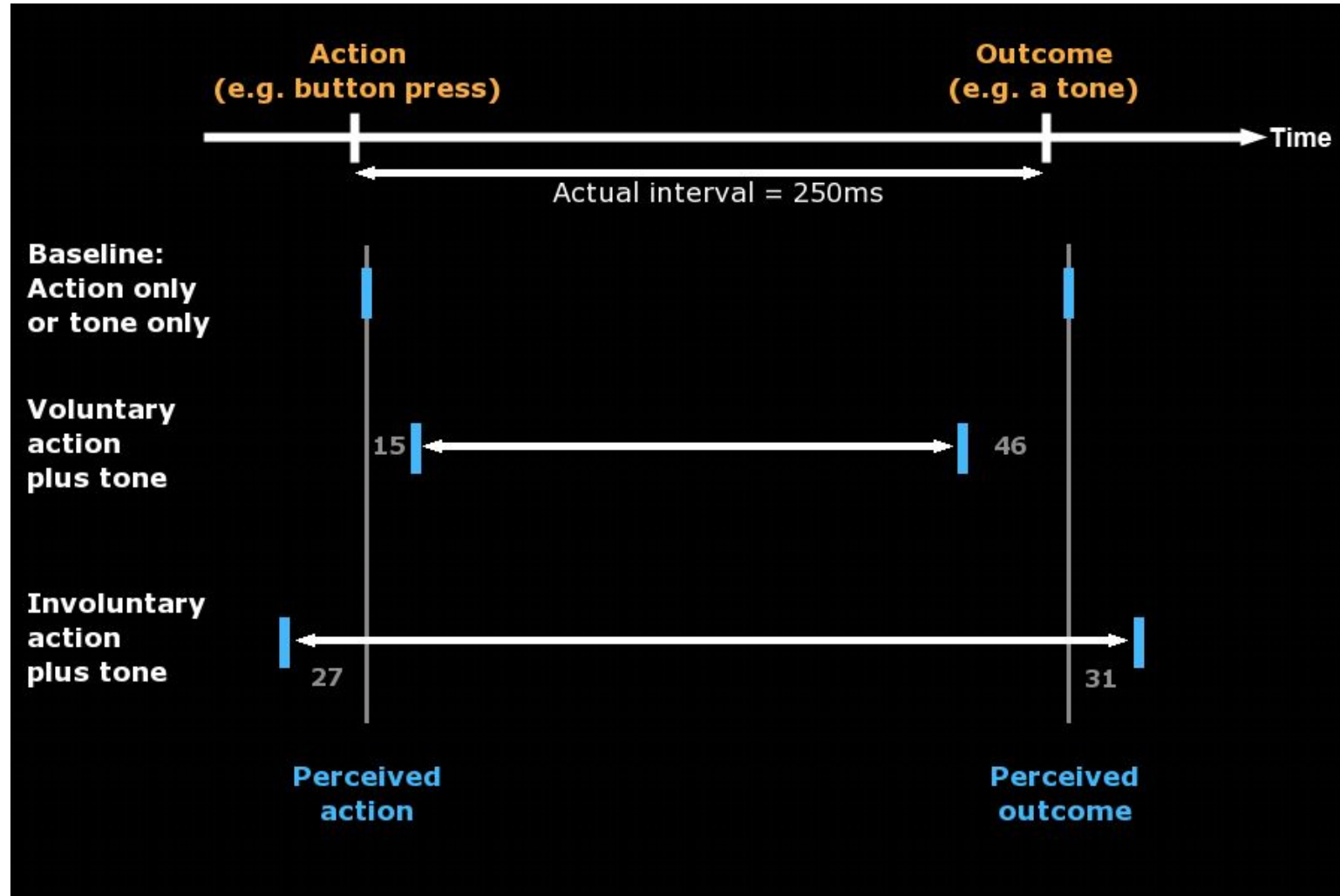
Haggard, P. & Tsakiris, M., *The Experience of Agency: Feelings, Judgments, and Responsibility*. Curr Dir Psychol Sci, 2009, 18(4) p.242-46.

Intentional binding



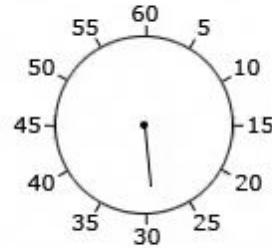
Haggard, P. & Tsakiris, M., *The Experience of Agency: Feelings, Judgments, and Responsibility*. Curr Dir Psychol Sci, 2009, 18(4) p.242-46.

Intentional binding



Haggard, P. & Tsakiris, M., *The Experience of Agency: Feelings, Judgments, and Responsibility*. Curr Dir Psychol Sci, 2009, 18(4) p.242-46.

The Libet clock method



- Approx. 100px in diameter.
- Shown at the centre of screen.
- Arm rotates once every 2560ms.

Strengths:

- Provides robust measures.
- Detailed breakdown of where binding occurs.

Weaknesses:

- Not suitable for visual tasks.
- Time consuming: 4 blocks of trials per condition.

Interval estimation

Participants estimate the time between their action and an outcome.

Strengths:

- Suitable for visual tasks.
- Less time consuming:
1 block of trials per condition.

Weaknesses:

- Less robust measure.
- No breakdown of where
binding occurs.

An experimental manipulation

- Skinput: appropriating the body as an input surface.
 - Harrison, Tan, & Morris. CHI 2010.



Images: © Chris Harrison

Experiment 1

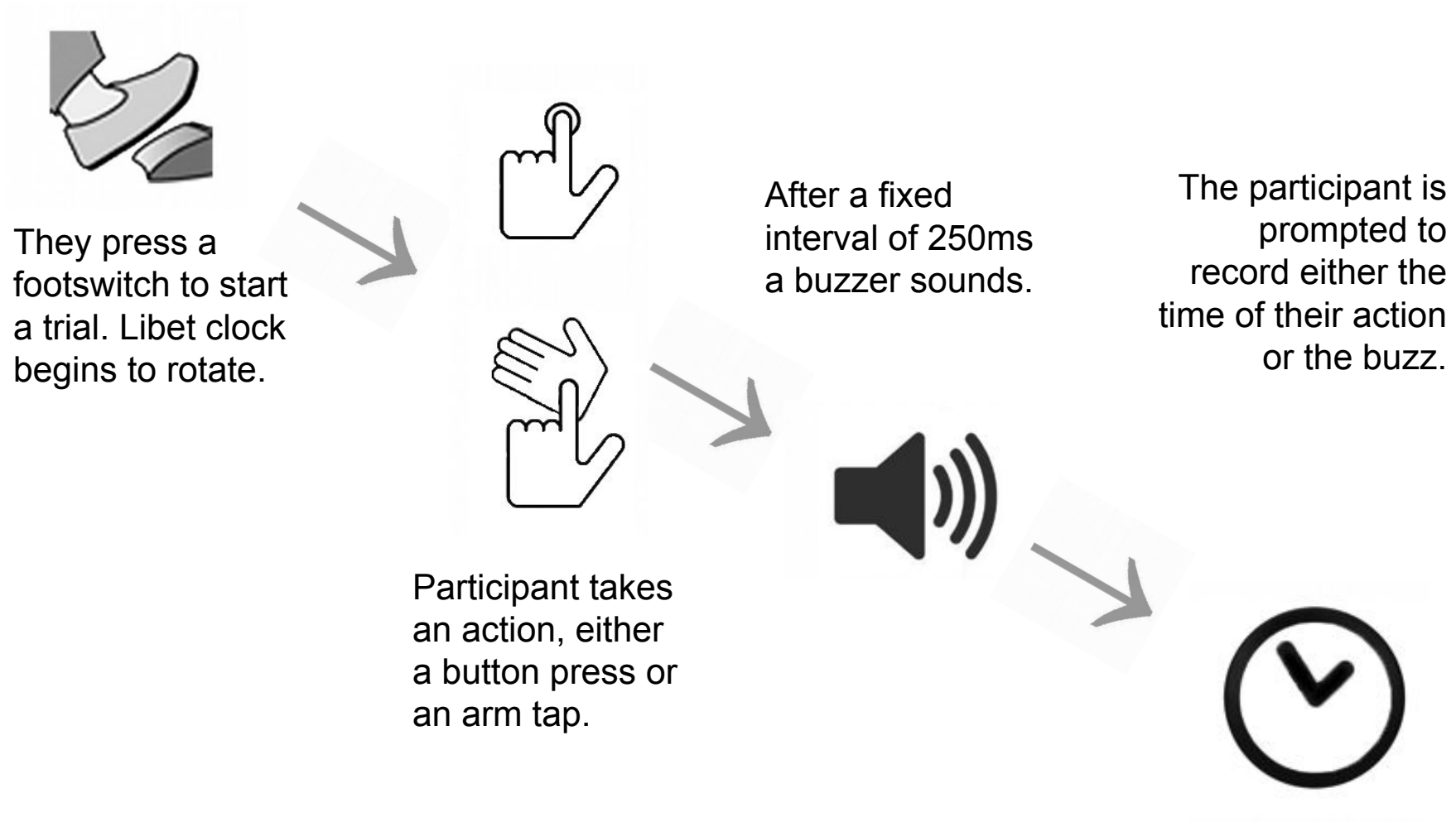
What's it like to be a button?

Do changes in the input modality of an action have an impact on the sense of agency?

Two input conditions: button and skin-based input.

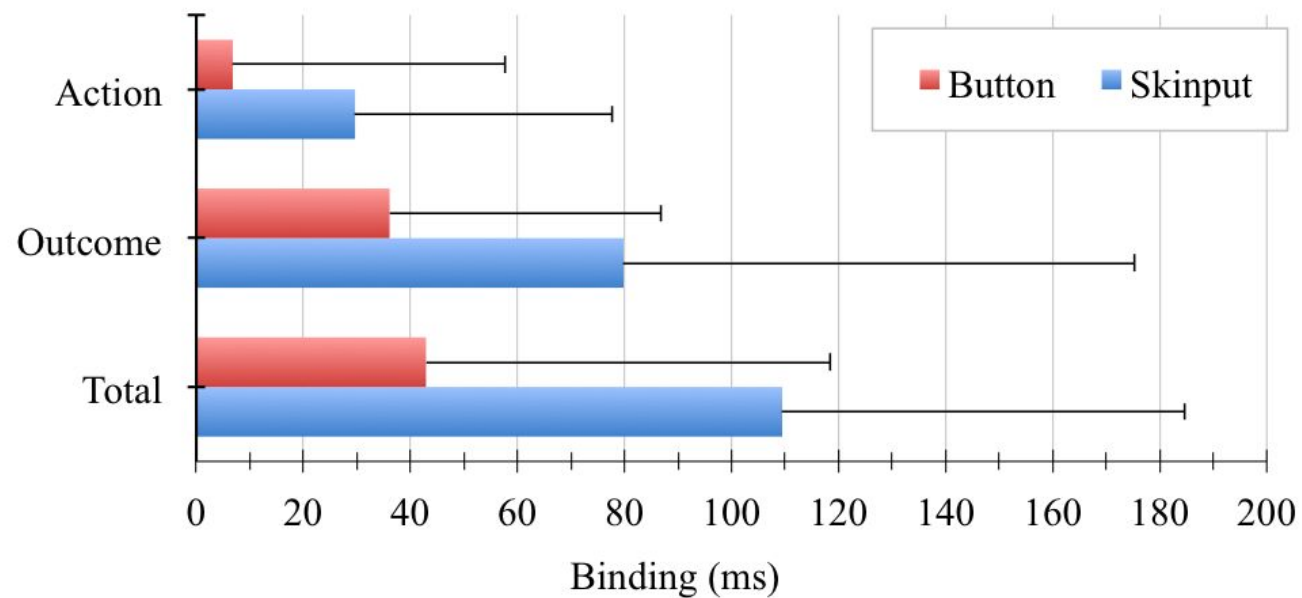


Procedure



Results

	Action binding	Outcome binding	Total binding
Button	6.81ms (45.6ms)	36.11ms (45.46ms)	42.92ms (67.43ms)
Skin-based	29.66ms (42.84ms)	79.82ms (91.23ms)	109.47ms (74.54ms)



**$t(18) = 4.05,$
 $p < 0.01$**

Observations

- Yes, changes in the input modality can have an impact on the experience of agency.
 - Intention binding is a useful metric for design research:
 - It can be used it to compare and refine input techniques.
 - Compare experiences for a given input technique when other conditions of the interactions change.
- For future research
 - What kind of intelligent interfaces might use the skin as an input surface?
 - Body-area networking is already a focus of commercial research
 - ‘Transhuman’ character Bethany in Russell T Davies’ BBC series *Years and Years*

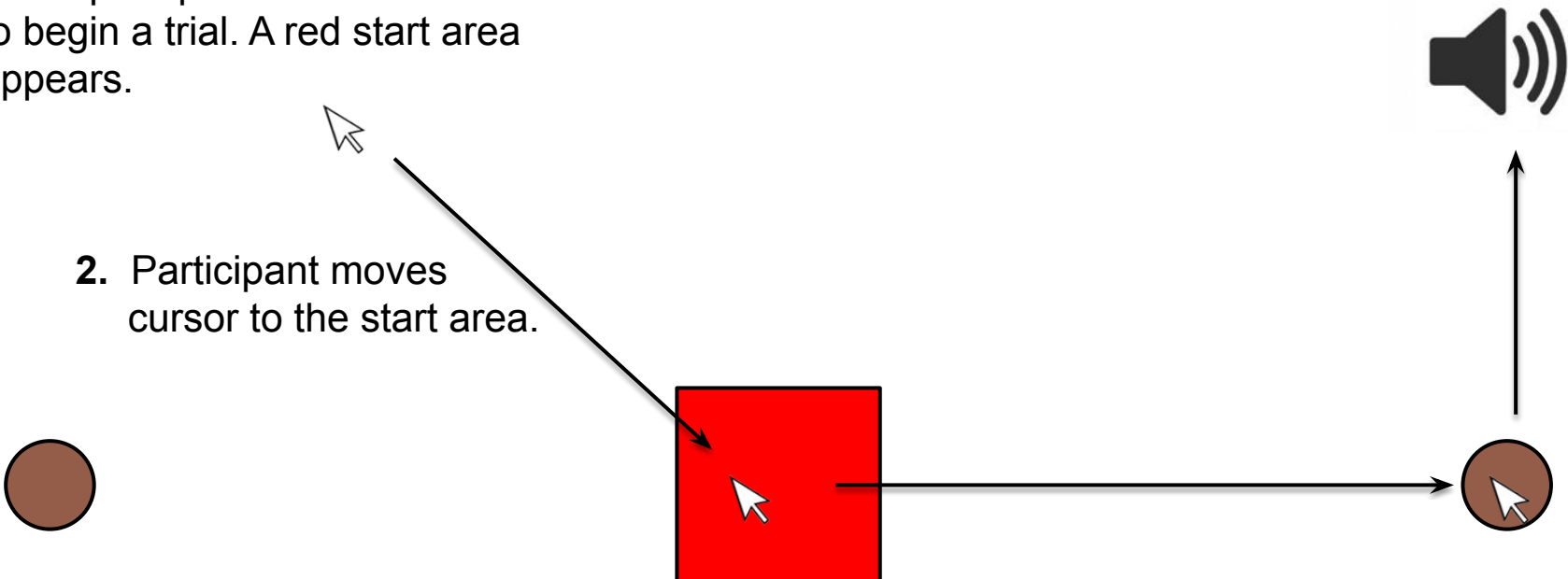


Experiment 2

Intelligent interfaces:

What happens when a computer helps out?

Procedure

1. Participant presses a footswitch to begin a trial. A red start area appears.
 2. Participant moves cursor to the start area.
 3. Having waited in start area for 1500ms two green targets appear.
 4. Participant chooses a target and moves cursor to hit it as quickly and as accurately as possible.
 5. Hitting target causes a buzz – with a random time interval.
The participant is prompted to estimate the interval between hitting the target and the buzz.
- 
- The diagram illustrates the experimental procedure with the following visual elements and flow:
- A brown circle on the left represents a target.
 - A red square in the center represents the start area, with a mouse cursor icon pointing to it.
 - A brown circle on the right represents another target, with a mouse cursor icon pointing to it.
 - A speaker icon at the top right indicates the buzz.
 - Arrows show the sequence: from the red square to the left target, from the red square to the right target, and from the right target to the speaker.

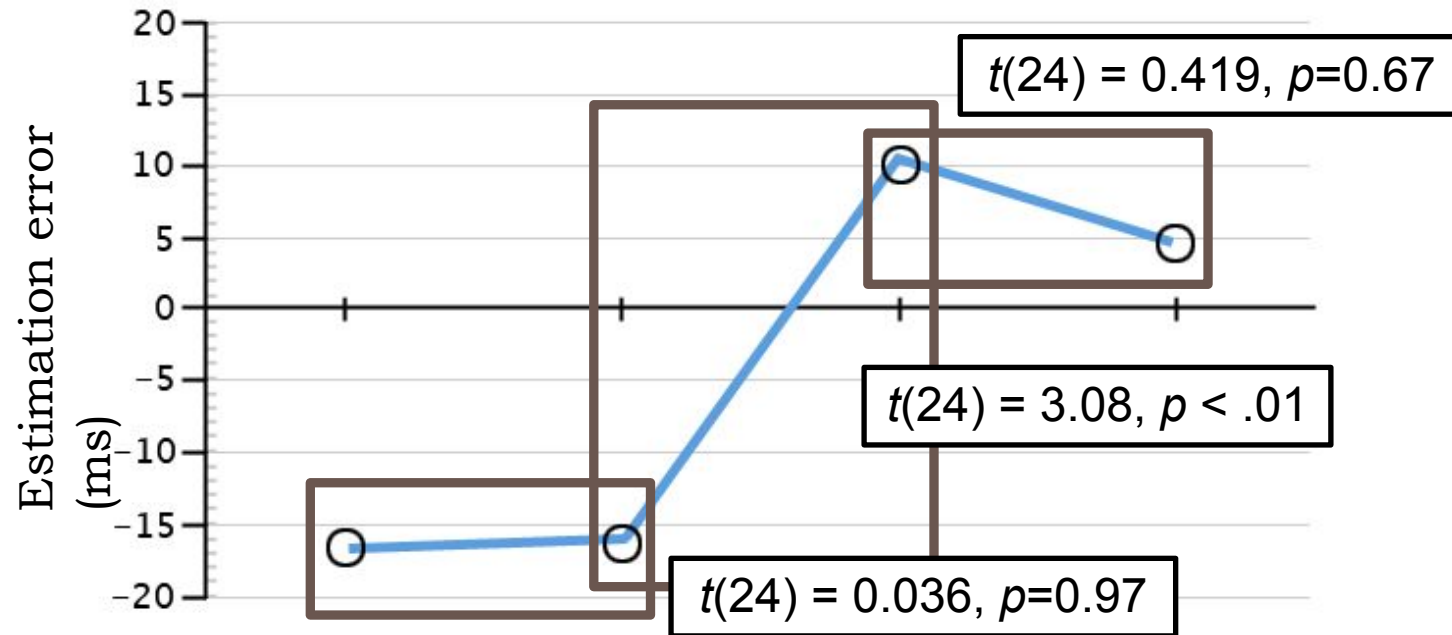
Experiment design

- Treatment: the assistance algorithm has the effect of adding “gravity” to targets.
 - Four levels of assistance: none, mild, medium, high.
- Within subject design, with:
 - 1 block of trials for each assistance level
 - 36 trials per block.
 - 24 participants.
- The order of the assistance level blocks was counter-balanced across participants.

Results

	No assistance	Mild assistance	Medium assistance	High assistance
Estimation error	-16.78ms (70.70ms)	-16.32ms (82.03ms)	9.93ms (85.92ms)	4.53ms (79.12ms)

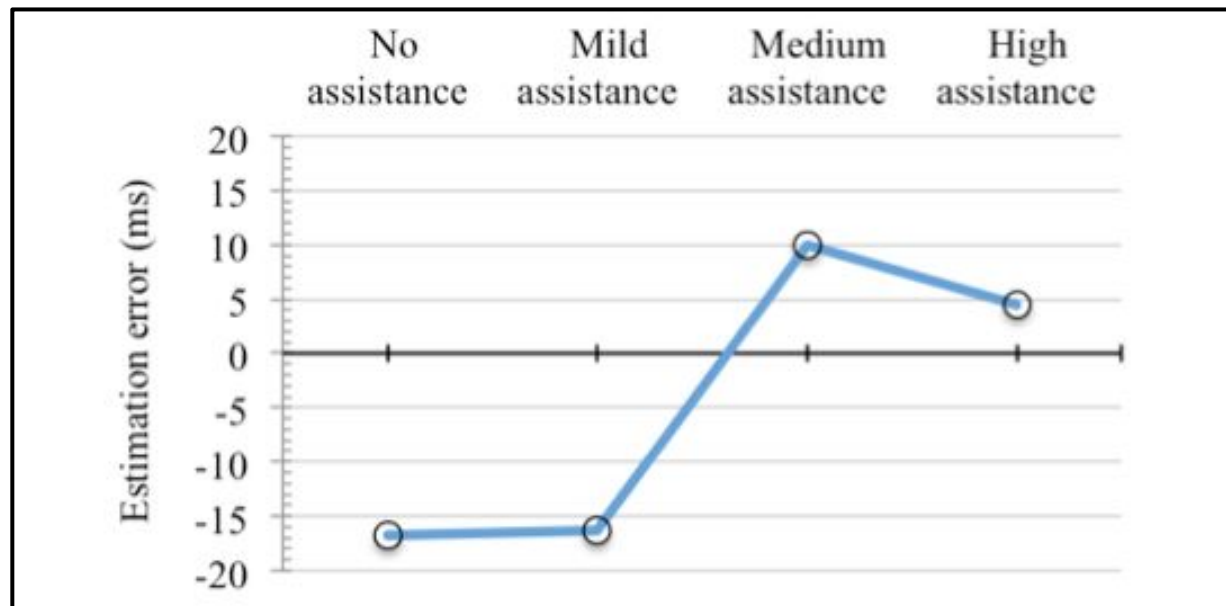
Greater intentional
binding, greater
sense of agency



Repeated measure ANOVA: $F(3,69) = 2.74, p=0.05$

Observations

- Up to a point, the computer gave assistance, but people retain a sense of agency.
- Beyond a certain point people experience a loss in sense of agency.
- This technique could provide an experimental means of mapping the personal agency characteristics of intelligent input techniques.



Experiment 3

Enhancing agency through timing in
“conversational” interfaces

Enhancing agency through timing in “conversational interfaces”

- Christine Yu Guo (2021)
 - Perception of Rhythmic Agency for Conversational Labeling (HCI Journal)
- Series of mixed-initiative experiments, comparing 4 treatments:
 - Computer-Arrhythmic (CA)
 - the Computer takes the initiative using aperiodic time intervals.
 - Computer-sets-Rhythm (CR)
 - the Computer takes the initiative using rhythmic timing.
 - User-followed-by-Computer (UC)
 - the User takes the initiative, and the Computer follows
 - *the idea is to imitate conversational entrainment between humans.*
 - User-sets-Rhythm (UR)
 - the User takes the initiative, setting their own rhythm – *today's industry standard*

Results

- ▮ *Predictable rhythms produce intentional binding:*
 - ▮ $CA < CR < UC < UR$
- ▮ *Predictable rhythms give sense of control:*
 - ▮ $CA < CR < UC < UR$
- ▮ *Predictable rhythms reduce task stress (NASA TLX scales)*
 - ▮ $CA > CR > UC > UR$
- ▮ *Predictable rhythms increase confidence of task success (NASA TLX scales)*
 - ▮ $CA < CR < UC < UR$

Overall conclusions

- Changes in the input modality and in levels of assistance can have a significant impact on users' experience of personal agency.
- Intentional binding can provide an implicit metric for probing and mapping experiences of agency.
- This metric can be applied in a wide range of design contexts. E.g.:
 - Comparison and refinement of different interfaces and assistance techniques.
 - Investigating the impact of uncertainty or different types of feedback.
 - Comparisons of user groups, e.g. different age groups, people experiencing mental health difficulties.
 - Enhancing mutuality in “conversational” mixed initiative



Design for control

Case Study: Coda (cf Reinforcement Learning from Human Feedback)

- Mixed initiative interface developed for Africa's Voices Foundation
 - <http://explain.avf.world/>
 - <http://www.africasvoices.org/ideas/newsblog/introducing-our-latest-analysis-tool-coda/>



- Guo explored effects of conversational rhythm when using this style of interface
 - Note that RLHF is *not* an HCAI approach!

Data receipt and coding

Now imagine you are a researcher at Africa's Voices Foundation and you have received text message responses from radio show listeners in Somalia. Your task is to label these data based on some of the emerging themes.

You label these data using a programme called CODA that uses artificial intelligence not to automate data analysis but to augment human ability to make sense of large sets of data.

Have a go at labelling some of the messages. 

Coda v2
Code by

scheme
 message

 Continuous sorting

↓ Next page

 oded ☒
 Keen Wombat

Seq	Message	Yes No	Reason
		Scheme-026f0a6f46e9	Scheme-38c75cbf1a05
51	yes i'm at risk due to bad weather	<input type="checkbox"/> <div></div>	<input type="checkbox"/> <div></div>
52	yes, family at risk because we don't have clean water	<input type="checkbox"/> <div></div>	<input type="checkbox"/> <div></div>
53	yes, we live in an overcrowded camp	<input type="checkbox"/> <div></div>	<input type="checkbox"/> <div></div>
54	yes. the statistics data shows there is an increasing potential of this risk.	<input type="checkbox"/> <div></div>	<input type="checkbox"/> <div></div>
59	yes, because there is no clean watwe	<input type="checkbox"/> <div></div>	<input type="checkbox"/> <div></div>

Mixed-initiative Coda concept

The interface displays a list of tweets with their IDs, text, and associated themes. A vertical bar on the left shows a stack of colored bars corresponding to the themes. A configuration panel on the right allows users to manage the themes and their shortcuts.

ID	Data	Theme
34	Yikes! Rape will burn temporary classes again. They needed to pay for these.	Protest
62	The DA is the only political party that handled the #Mabellansen situation correctly, without using it for politics.	Racism (judicial system)
66	This #Mabellansen lady is an entire judge. A whole one. And this is her logic. Yhoooooaaaaa:	Racism (judicial system)
87	A message for all rape survivors out there: We believe you #UCTSpeaksUp #EndRapeCulture #UCTSurvivors	Rape Culture
98	#Mabellansen's black adopted daughter said the Judge is not a #racistMNC in #SouthAfrica has #racism syndrome.	Racism
73	By attributing rape to a specific designation, you are effectively telling white women that this will not happen if y're white. #Mabellansen	Racism (judicial system)
78	Black need to stop seeking White acceptance. They'll never get it, not with #Mabellansen minds. No legislation will change them. Move on.	Racism (judicial system)
80	What's happened to Khumalo when said he wanted to cleanse the country of all white people. See Hitler did with the Jews. #Mabellansen	Racism (judicial system)
81	There are many people in positions of power who have no place there and #Mabellansen is one of them.	Racism (judicial system)
82	By the logic of #Mabellansen then white culture is racism...	Racism (judicial system)
104	#FridayStandIn #Mabellansen NOT a killer Black Hate Greed Injustice Corruption Hypocrisy is deadly	Racism (judicial system)
107	UCT Speaks Up LOUDLY! Rape Survivors #RUPreferenceList	Rape Culture

Configuration Panel:

Code Scheme Name: overall theme

Code	Shortcut
Protest	type shortcut key...
Racism (judicial system)	type shortcut key...
Rape Culture	type shortcut key...
Racism	type shortcut key...

Code Color: #e4ec38

Regex:

Words:

Simplified version for controlled experimental study

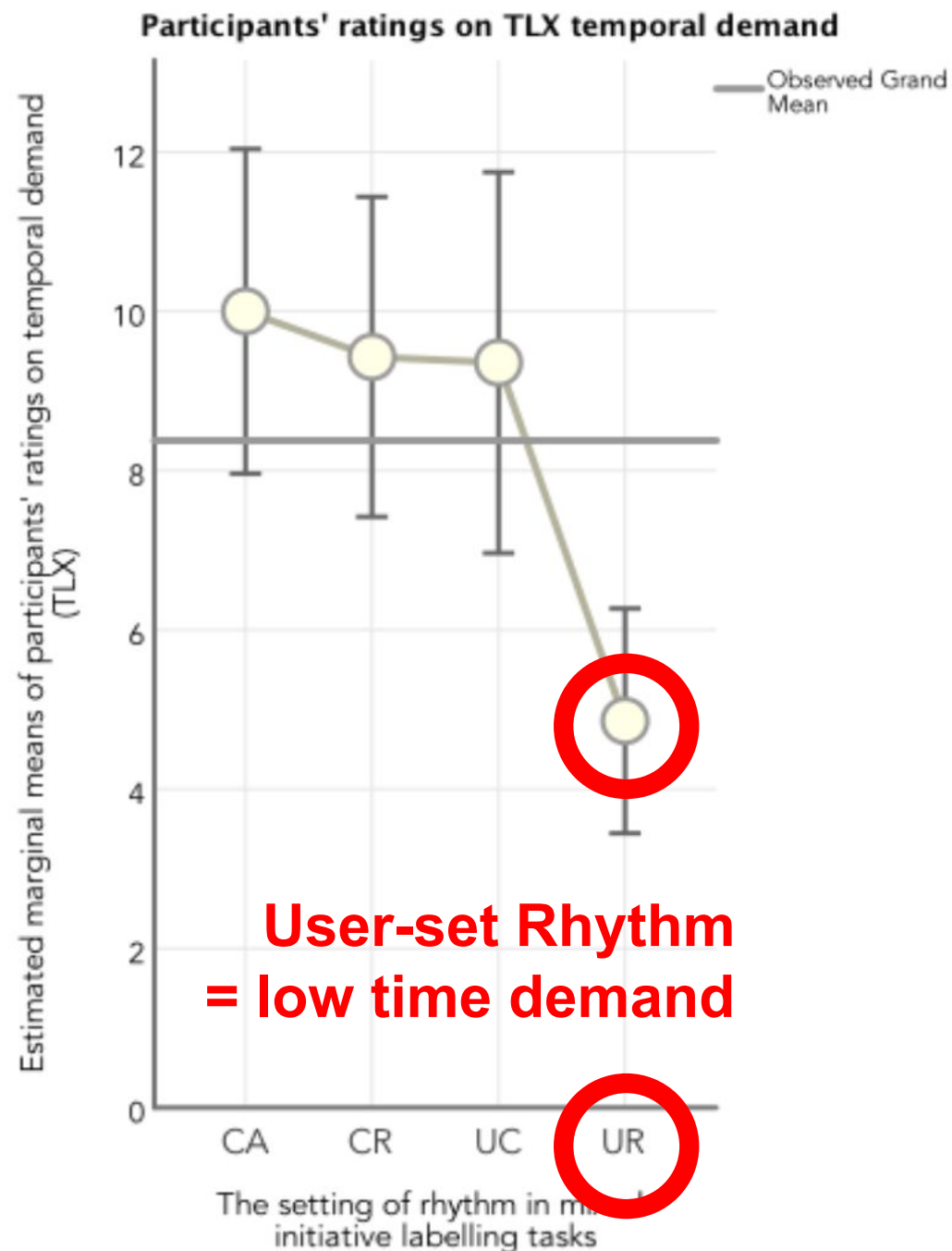
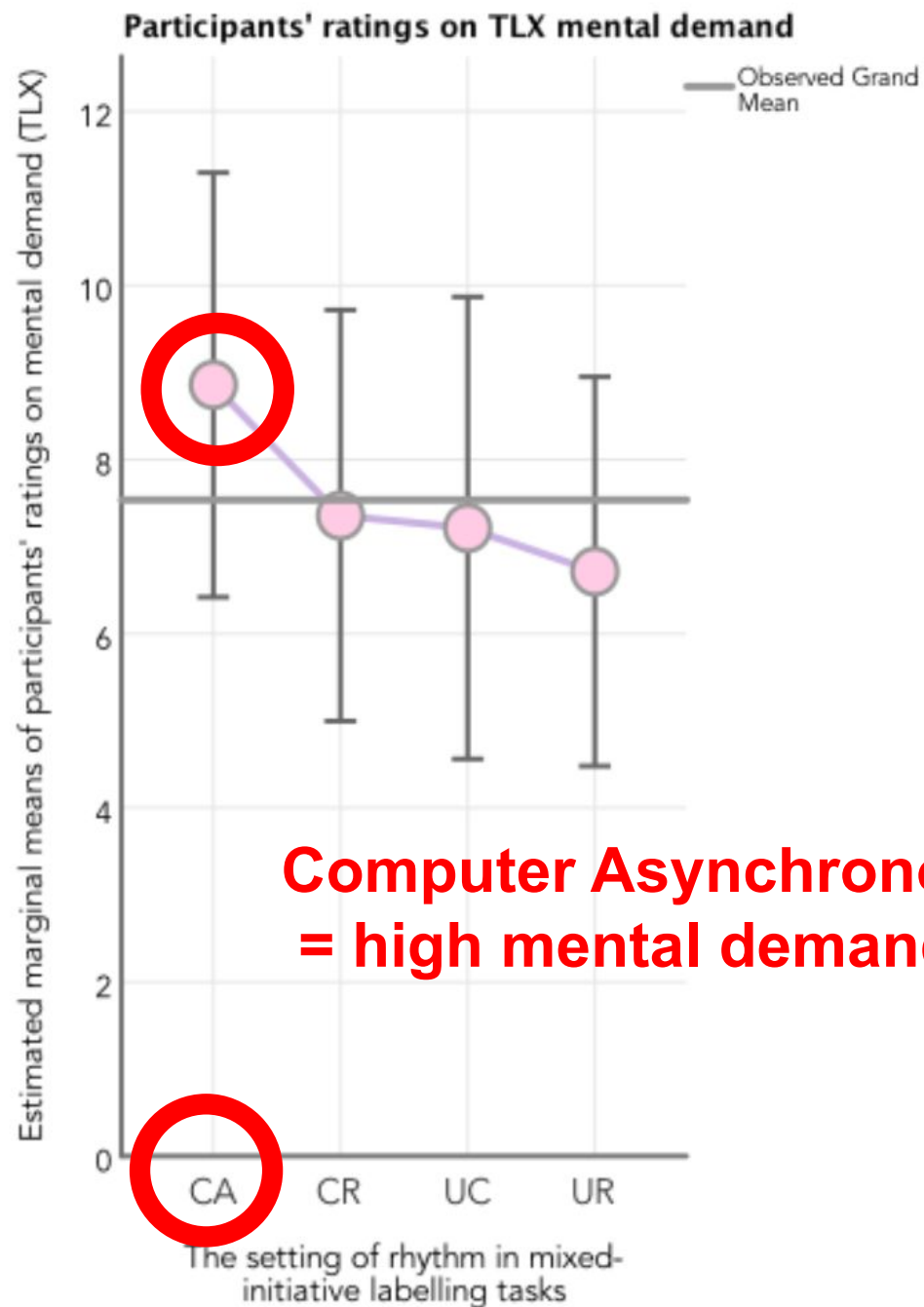
No.	Time	Content	Computer's Judgement	correct?	wrong?
41	1/2/2017 6:02:18 PM	Can I ask if you have the Game of Scones baking tray?	It is about delivery.	Correct	Wrong
42	1/3/2017 5:50:11 PM	When could you deliver the Kallax shelves to my office?	It is about delivery.	Correct	Wrong
43	1/4/2017 5:02:12 AM	Just checking if this saucepan works on an induction oven?	It is NOT about delivery.	Correct	Wrong

Each participant makes 30 decisions with each of the 4 different interfaces (CA/CR/UC/UR)



Confusion matrix of experimental stimuli – user should correct these

		Label presented by the system		Totals
		Delivery	Not Delivery	
Actual content of the text	Delivery	5	5	10
	Not Delivery	10	10	20
Totals		15	15	30



Confusion matrix of user labelling performance

		User judgement of the label	
		Wrong	Correct
Accuracy of system label	Wrong	211 / 215 / 216 / 215	14 / 10 / 9 / 10
	Correct	3 / 8 / 3 / 10	222 / 217 / 222 / 215
Totals		214 / 223 / 219 / 225	236 / 227 / 231 / 225

Table 3 - confusion matrix with breakdown by condition CA / CR / UC / UR

User-set Rhythm results in less 'mutual respect' for system judgments, more errors from false positive rejections