

Planning and Situating Actions: challenges for explanation and trustworthiness in autonomous systems

Prof Paul Luff

Work, Interaction and Technology
King's Business School

THuMP

trust in
human-machine
partnerships

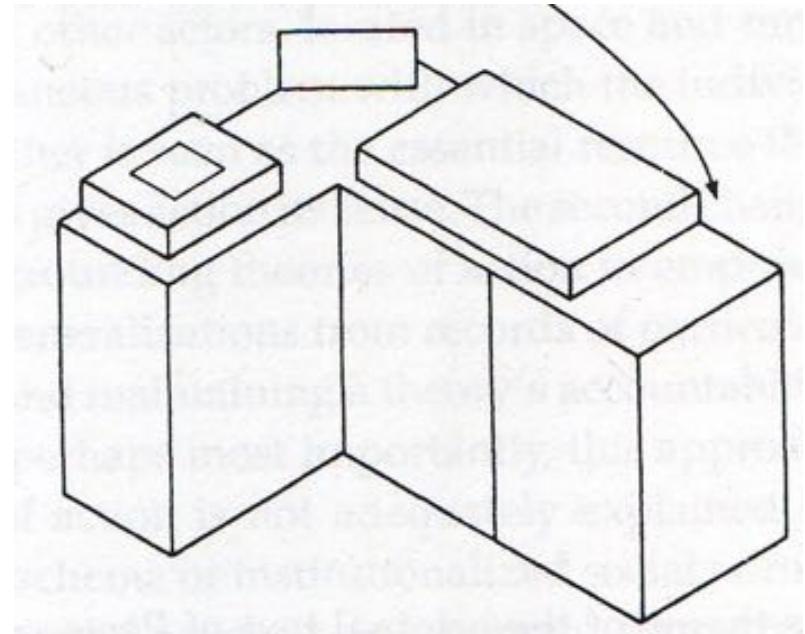


UKRI
Trustworthy
Autonomous
Systems Hub

KING'S
College
LONDON

Background: Trust, Explanation and Autonomy

- interdisciplinary collaboration
- concepts become problematic - unpacking and refining
- critiques and challenges
- Suchman (1987): intelligent photocopier (physical system) - plans, goals
- practical implications
- critique of concepts: vs cognitive science
- methodological consequences - different kind of studies in HCI



Suchman (1987/2007) Human-Machine Reconfigurations: Plans and Situated Actions (2nd Edition), Cambridge University Press

Background: Work, Interaction & Technology

- video based studies of social interaction - informed by ethnomethodology and conversation analysis
- organisations: tacit skills and practices: talk, gesture and material conduct
- settings: surgery, control rooms, trading rooms, news rooms, dentistry, optometry, auctions, education, medical practice, architecture, design, construction, markets, museums and galleries, auctions...



Outline

- ⌚ Approach and orientation - recent studies related to material objects and ‘physical systems’ (robots)
- ⌚ Workplace study - surgery and the passing of instruments (ongoing)
- ⌚ Quasi-naturalistic study - THuMP - trust, planning and physical systems (starting)
- ⌚ Workplace Study and Quasi-naturalistic Study: TAS - trustworthy autonomous systems, teamwork and robots (preparing)
- ⌚ Considerations and Challenges

1. Workplace Study: Surgical operations - physical objects

- surgery as complex environments - embodied collaboration
- surgical instruments: variety of artefacts: familiar and unfamiliar appearances
- one recurrent activity - the **passing of instruments** - a smooth exchange
- surgeons require timely and relevant access to an instrument - provided by staff with a specific role 'scrub nurses'
- rarely explicit verbal request or instruction
- ***right time, right place, right way***



Heath and Luff (2018). Exchanging implements: the micro-materialities of multidisciplinary work in the operating theatre. *Sociology of Health and Illness*, Special Issue on Materialities of Care. 40, 2 (297-313)

Collaborative exchange of instruments

- ⌚ preparatory actions: positioning of instruments
- ⌚ exchange shaped with regard to contingencies of the operation:
 - ⌚ the particular procedure
 - ⌚ ‘disruptions’ to the procedure

progressive articulation

- ⌚ subtle differences in its accomplishment progressive transformation of the activity
- ⌚ contributions to passing by both participants: prospectively envisaged and finely tuned

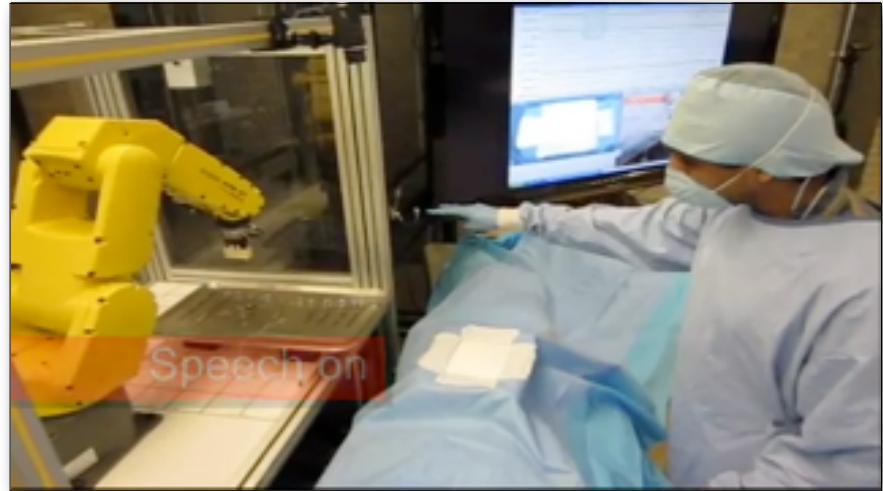


Implications: Trust, Organisation and Interaction

- relations between professions - scrub nurse and 'support'
- collaboration and tacit interaction - produced in interaction
- 'knowledge', 'familiarity' and 'expertise'

robotic scrub nurse

- safety in surgery - avoiding errors
- mischaracterising the activity - two actions ('place' and 'grab') not one continuous activity



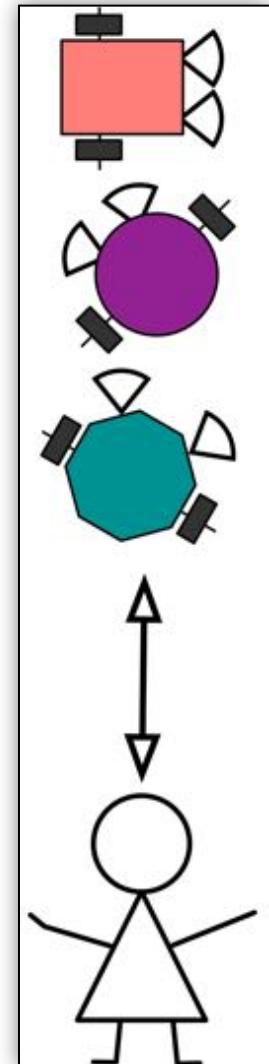
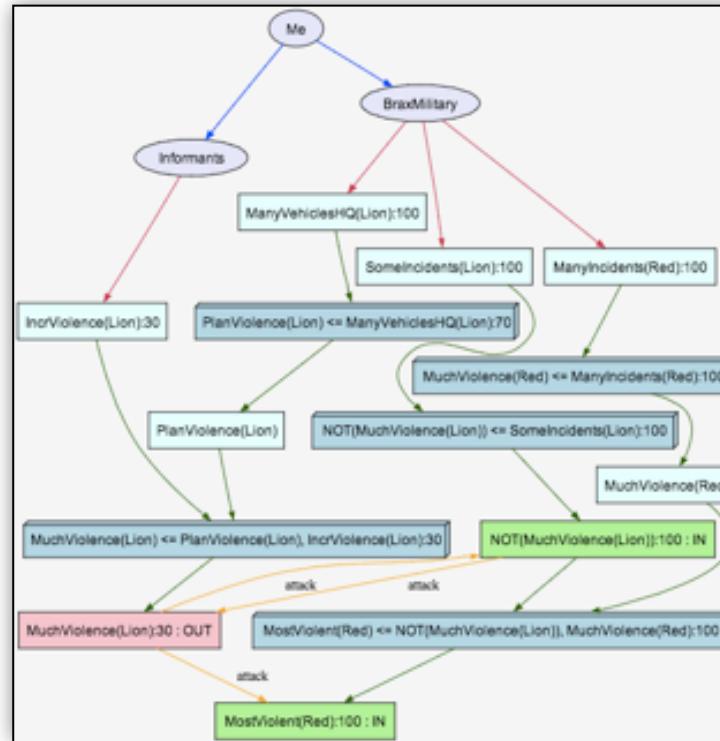
Jacob, Li, Akingba and Wachs (2011) Gestonurse: A robotic surgical nurse for handling surgical instruments in the operating room. *Journal of Robotic Surgery*, 6(1), 53–63.

Collaboration with a robotic scrub nurse. *Communications of the ACM*, 56(5), 68–75.

2. Simulation Study: Trust in Human-Machine Partnerships



- Explainable Artificial Intelligence **Planning** (XAIP) systems (non-physical)
- technical issues: contrastive explanations, argumentation, visualisation, provenance
- social, legal and ethical challenges
- complex plans in organisations and robots



Canal, G. Huynh, D., Krivic, S. Mahesar, Q-A, Borgo R., Coles, A., Moreau, L., Patel, M., Luff, P., Drake, A., Keller, P. & Parsons, S. (2020) Trust in Human Machine Partnerships, WebSci'20 Workshop: Explanations for AI: Computable or Not?, Southampton, July 7–8, 2020

Robot Simulation Studies

- ⌚ pilot for ‘real’ robot studies (Toyota HSR)
- ⌚ programme of studies
 - verbalisation, argumentation, visualisations, provenance
- ⌚ public engagement
- ⌚ robot intern study



Robot Intern Study: verbalisations



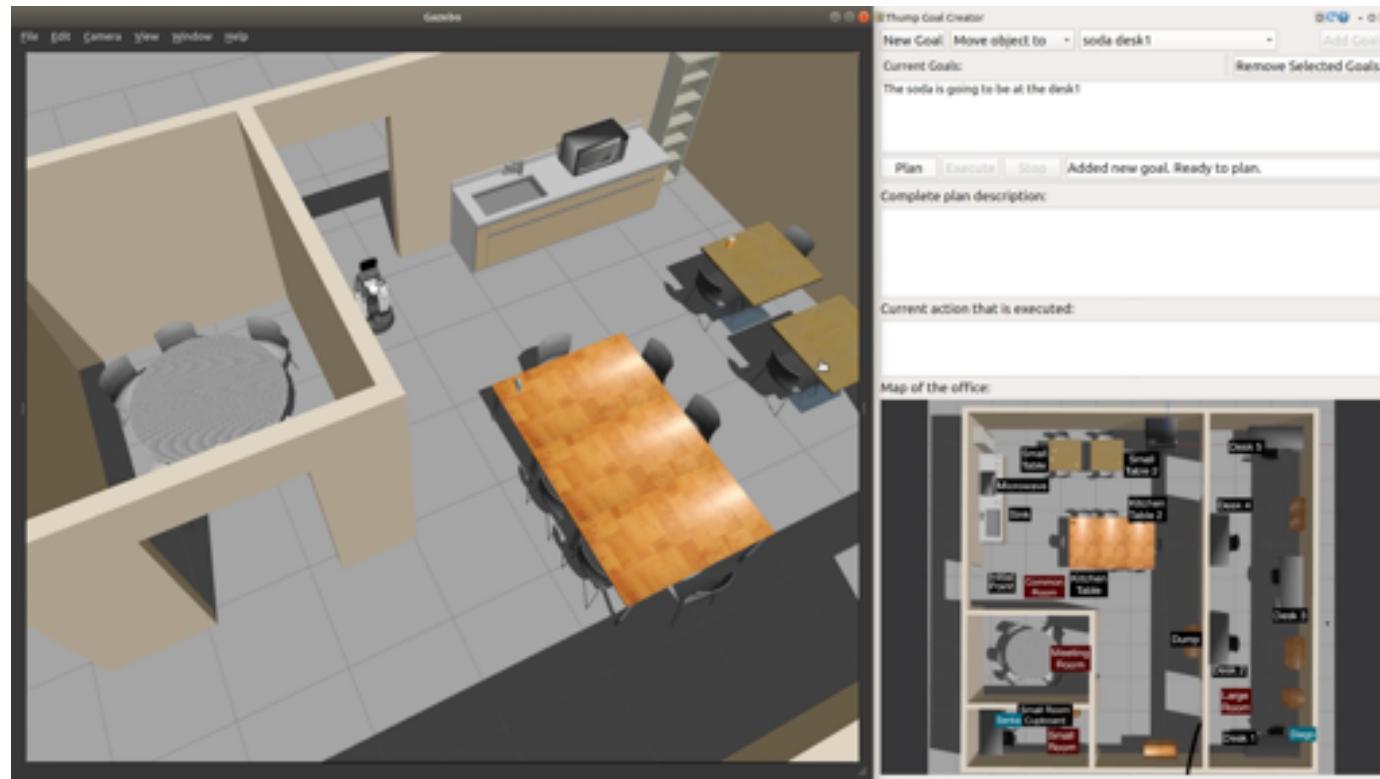
pre, during and post



Robot Intern Study

THuMP

- ⌚ observational study of pairs of users (+interviews)
 - ⌚ Analysis of actions and interventions on the system and dialogue between users
 - ⌚ Explanations in context
 - ⌚ Explanations in interaction
 - ⌚ breaching experiments in trust - studies and simulations (WoZ) - trust issues



3. Workplaces and Simulation Study: Robot Assisted Surgery

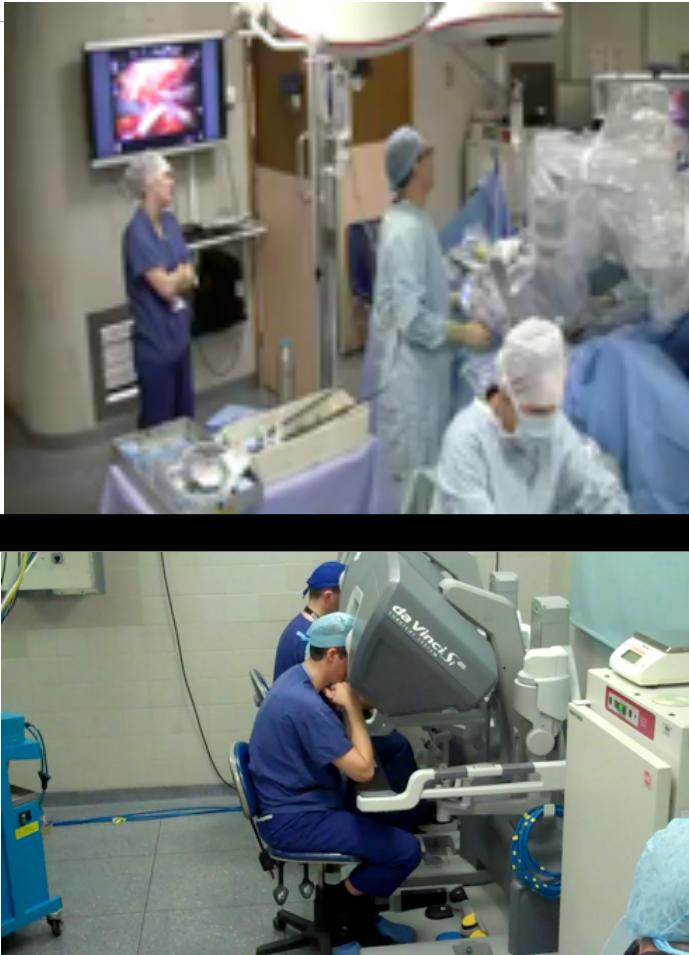
- ⌚ Trustworthy Autonomous Systems Hub (£13m)
 - distinctive kind of research funding
- ⌚ Projects in 'Trust', AI and robotics
- ⌚ domains include: healthcare, finance, autonomous vehicles, home, defence & security, creative industries
- ⌚ agile, adaptive and interdisciplinary approach
- ⌚ example studies: robotic surgery, financial decision-making, robot deep cleaning...



Robot Assisted Surgery



- robot and teamwork: situated and peripheral awareness, mediated collaboration
- autonomy and trust - by patients, carers, staff, etc - direct and indirect
- Trust in teams, trust of teams
- Definitions of autonomy
- problems - trust established in interaction and as part of work
- breaching experiments: simulations, communication and repair in teams



Legacy of Suchman: workplace studies and technology

- ⌚ range of studies: focus on situated - interaction, activities and practice
 - ⌚ research programmes vs research projects
 - ⌚ substantive - informing requirements, design, assessment and deployment
- ⌚ methodological - the quasi-naturalistic experiment: exploring technological innovations and social interaction
 - ⌚ informing technological development and informing social science
 - ⌚ disrupting the everyday - making the ordinary strange
- ⌚ analytic - materiality, trust, agency
 - ⌚ critique of concepts - when not grounded in the real(everyday) world - 'trust', 'goal', 'plan'



Concerns with ‘trust’, ‘explanation’, ‘autonomy’

Concerns about concepts - danger of category mistakes (Ryle)

Concerns with taxonomies and ontologies - problem of relevance (Schegloff)

Concerns with abstraction and ‘toy problems’ - understanding actions in contexts, with organisations and context - practices, purposes, warrants, accountability, designs for different participants

Ryle, Gilbert (1949) *The Concept of Mind*: Hutchinson: London

Schegloff, Emanuel A. (1991), "Reflections on talk and social structure", In *Talk and Social Structure: Studies in Ethnomethodology and Conversation Analysis* (Deirdre Boden, Don H. Zimmerman, eds.), Cambridge, Polity Press, pp. 44–70.

Considerations and Challenges

- ⌚ **Explanations:** - from uni-directional - to being produced in interaction, context shaped and context renewing, **recipient design (conversation analysis)**
- ⌚ **Trust:** - from being a property - to a practice, implicit, how accessible through explicit means (questioning)?
- ⌚ **Physical Systems:** - not just talk and text - embodied explanations, visual and material conduct includes objects and artefacts (HRI)

Suggestions

**Complex concepts (e.g. trust, autonomy (agency) and explanation
contributions from ethnomethodology and conversation analysis**

- **Membership categorisation devices** - common sense (ordinary language) categories and context - how produced in context, how relevant to participants (the participants' perspective)
- **Sacks's gloss** - where problematic, find occupations where it is a concern and analyse them and their practices (explanation, trust, agency...)
- **Everyday methods** (Garfinkel) - moral conception of trust - 'trust' relevant when problematic - breaching experiments - what constitutes trust?

Sacks, Harvey (1992) Lectures in Conversation: Blackwell Oxford

Garfinkel, H. (1962) A conception of and experience with 'trust' as a condition of stable concerted actions. In Harvey, O.J. (ed) Motivation and Social Interaction. New York: The Ronald Press