An exploration of measurement in human-centred computing research

Sandy Gould

School of Computer Science and Informatics,
Cardiff University

goulds@cardiff.ac.uk — @sjjgo — sjjg.uk

Photo: Saeid Anvar on Pexels

Picking up an old thread…

What I'll talk about…

- Why it is worth spending time talking about measurement.
- A tour of some of my attempts to measure things.
- Disciplinary* norms for what to measure and how to measure it.
- Zeitgeisty influences on practice.

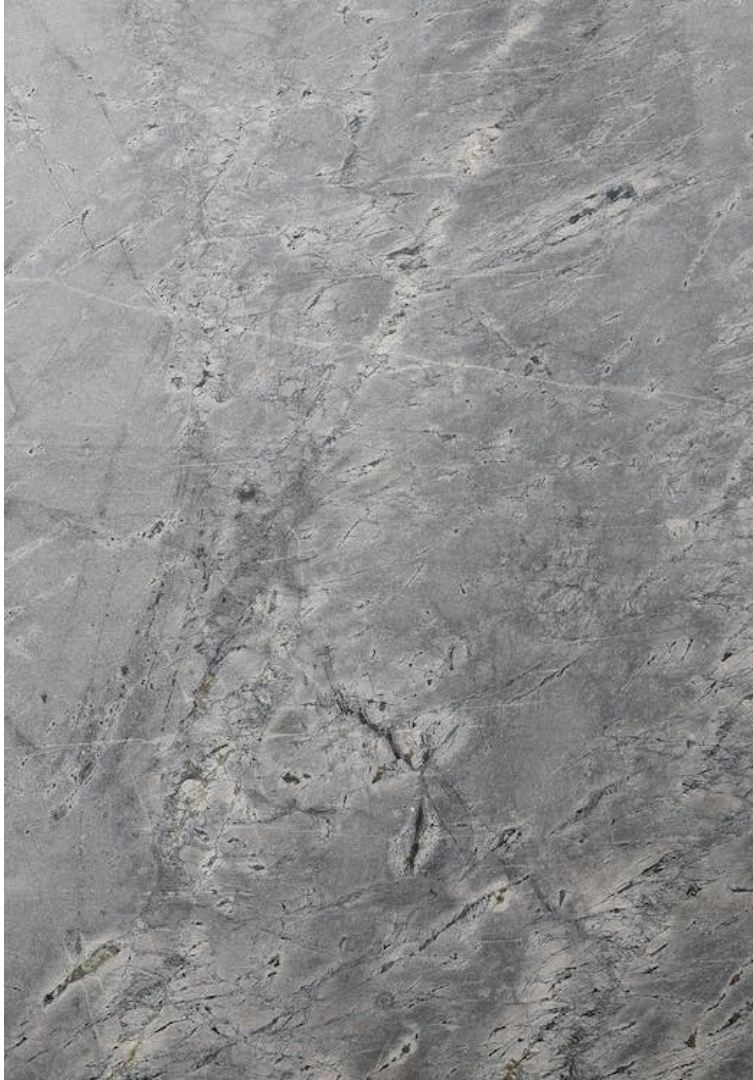
*See Reeves (2015) Human-computer interaction as science.

Measurement

- This talk is not about quantification.
- Interviews can be a form of measurement.
- I'm talking about all empirical research in human-centred computing.
- What we measure and how we measure significantly constitutes a contribution to knowledge.

Measurement

- Construct validity is essential to good empirical work.
- We are often trying to minimise the gap between what we *can* measure and what we really *want* to measure.
- But there a lot of constraints that wedge that gap open: ethical, logistical, epistemological, &c



A critique of measurement in my work

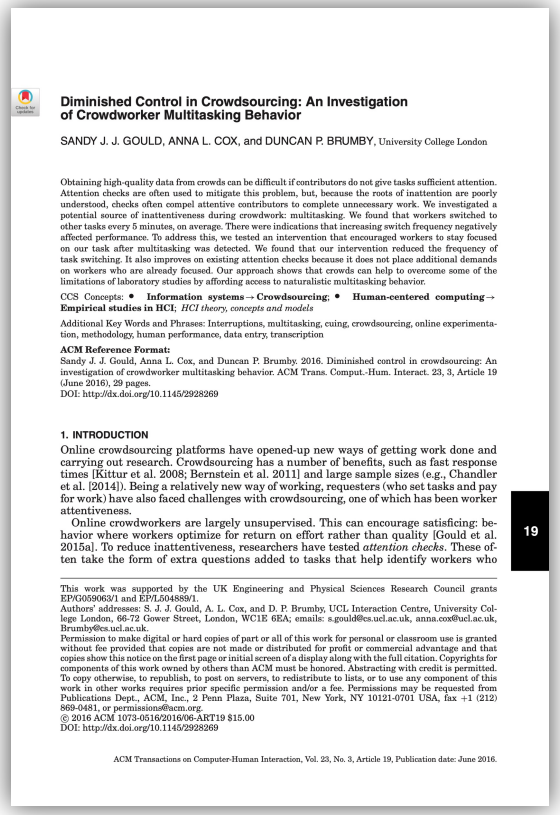
To make the discussion of constraints more salient, I will start by revisiting some of my work to think about what blocked achieving ‘better’ construct validity.

I’ve chosen work that I led on; I have greater insight into the decisions made, and it’s easier to poke and prod work that you took the biggest decisions about!

Example: Inferences about interruptions

Context:

- 2016 ToCHI paper
- How do people doing online crowdwork interleave their various tasks?
- Received July 2014; revised February 2016; accepted April 2016



Diminished Control in Crowdsourcing: An Investigation of Crowdworker Multitasking Behavior

SANDY J. J. GOULD, ANNA L. COX, and DUNCAN P. BRUMBY, University College London

Obtaining high-quality data from crowds can be difficult if contributors do not give tasks sufficient attention. Attention checks are often used to mitigate this problem, but, because the roots of inattention are poorly understood, checks often compel attentive contributors to complete unnecessary work. We investigated a potential source of inattentiveness during crowdwork: multitasking. We found that workers switched to other tasks every 5 minutes, on average. There were indications that increasing switch frequency negatively affected performance. To address this, we tested an intervention that encouraged workers to stay focused on our task after multitasking was detected. We found that our intervention reduced the frequency of task switching. It also improves on existing attention checks because it does not place additional demands on workers who are already focused. Our approach shows that crowds can help to overcome some of the limitations of laboratory studies by affording access to naturalistic multitasking behavior.

CCS Concepts: • **Information systems** → **Crowdsourcing**; • **Human-centered computing** → **Empirical studies in HCI**; *HCI theory, concepts and models*

Additional Key Words and Phrases: Interruptions, multitasking, cuing, crowdsourcing, online experimentation, methodology, human performance, data entry, transcription

ACM Reference Format:
Sandy J. J. Gould, Anna L. Cox, and Duncan P. Brumby. 2016. Diminished control in crowdsourcing: An investigation of crowdworker multitasking behavior. ACM Trans. Comput.-Hum. Interact. 23, 3, Article 19 (June 2016), 29 pages.
DOI: http://dx.doi.org/10.1145/2928269

1. INTRODUCTION

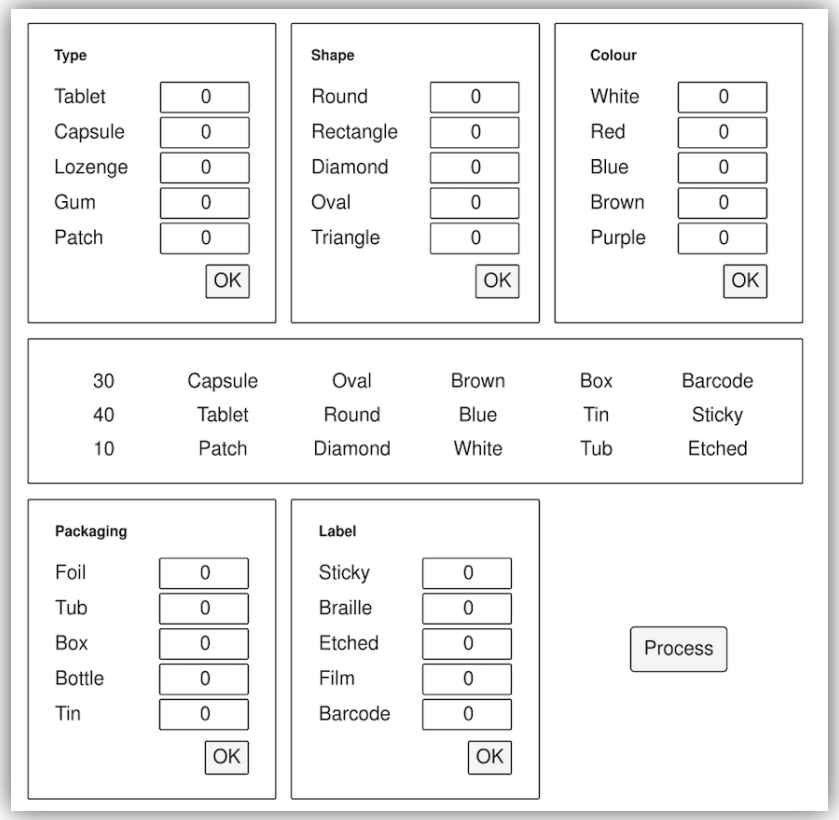
Online crowdsourcing platforms have opened-up new ways of getting work done and carrying out research. Crowdsourcing has a number of benefits, such as fast response times [Kittur et al. 2008; Bernstein et al. 2011] and large sample sizes (e.g., Chandler et al. [2014]). Being a relatively new way of working, requesters (who set tasks and pay for work) have also faced challenges with crowdsourcing, one of which has been worker attentiveness.

Online crowdworkers are largely unsupervised. This can encourage satisficing: behavior where workers optimize for return on effort rather than quality [Gould et al. 2015a]. To reduce inattentiveness, researchers have tested *attention checks*. These often take the form of extra questions added to tasks that help identify workers who

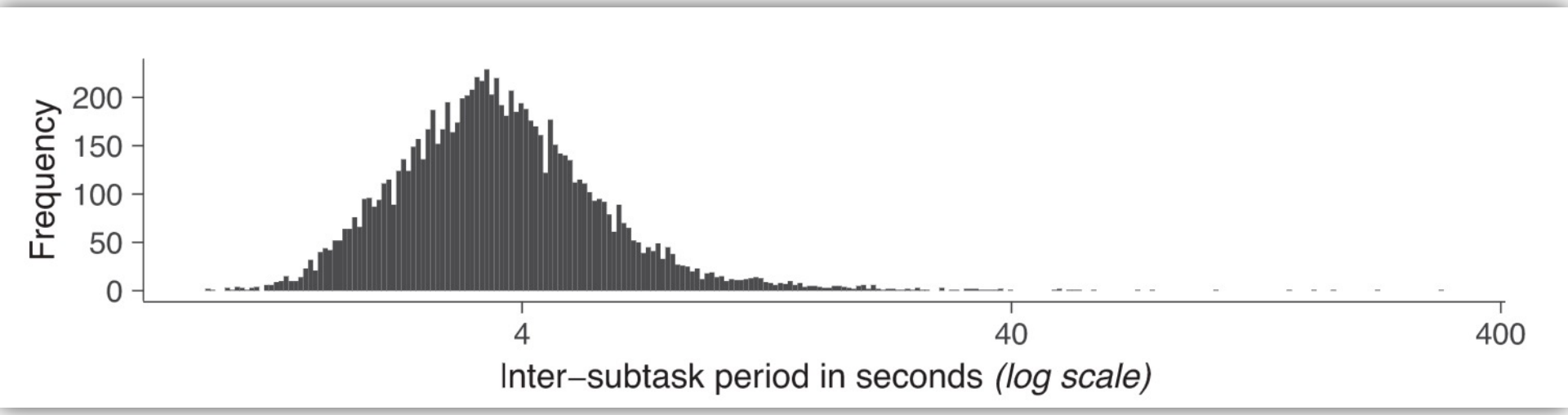
This work was supported by the UK Engineering and Physical Sciences Research Council grants EP/G059063/1 and EP/L504889/1.
Authors' addresses: S. J. J. Gould, A. L. Cox, and D. P. Brumby, UCL Interaction Centre, University College London, 66-72 Gower Street, London, WC1E 6EA; emails: s.gould@cs.ucl.ac.uk, anna.cox@ucl.ac.uk, Brumby@cs.ucl.ac.uk.

DOI: http://dx.doi.org/10.1145/2928269

Example: Inferences about interruptions



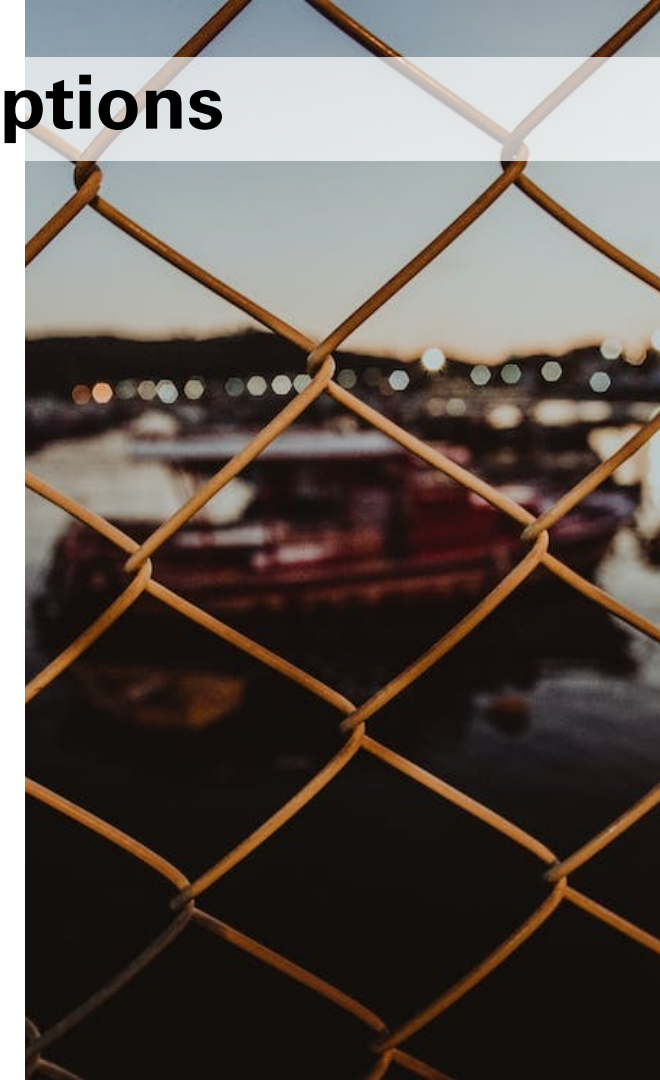
Example: Inferences about interruptions



Example: Inferences about interruptions

Constraints:

- Ethical; obligation not to collect to excess
- Resource; time, done during PhD
- Logistic; challenges of deploying outside the browser
- Epistemological; still couldn't have 'seen' beyond the computer



Example: Inferences about interruptions

Possible alternative:

- Build a browser plugin (see, e.g., Toxtli et al. 2021)
- Logistical challenges of deployment against more granular data
- Stuck inside the browser though. Do you build a desktop application? Do you start collecting webcam data?

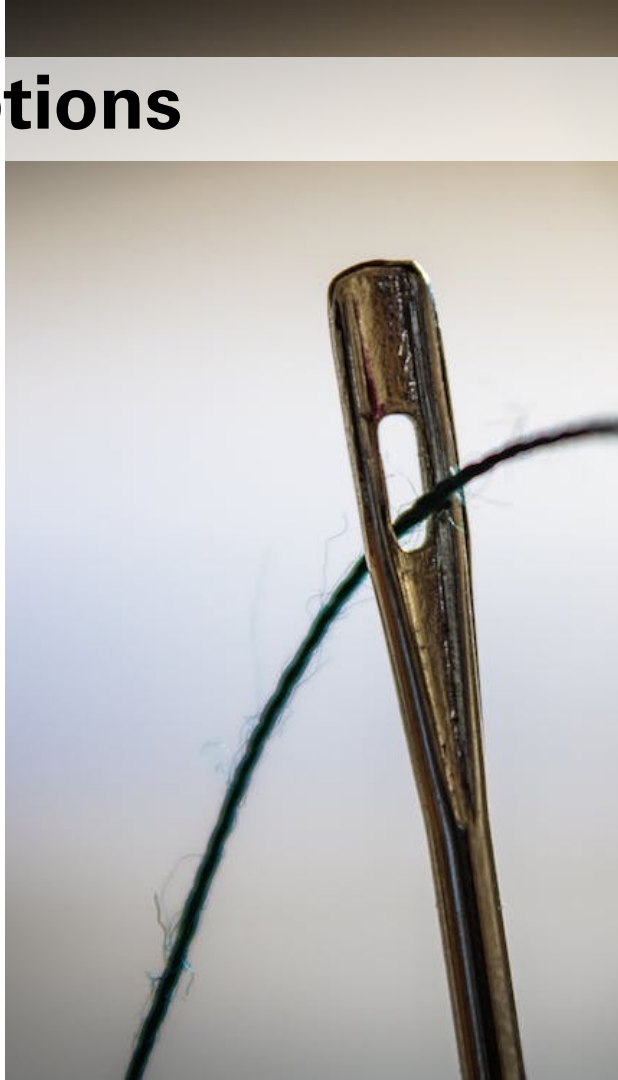
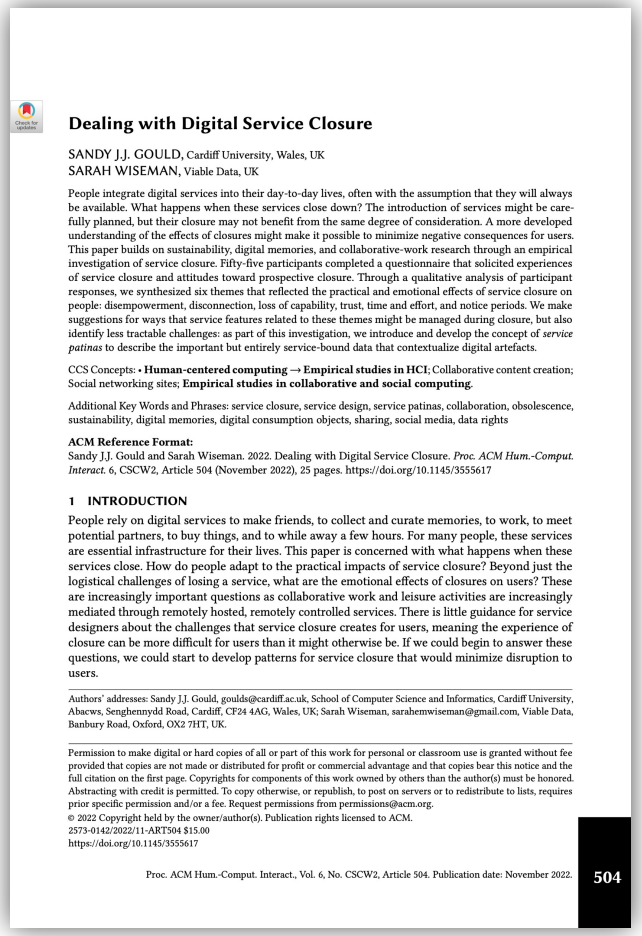


Photo: David Riaño Cortéson Pexels

Example: service closure

Context:

- 2022 CSCW paper
- What does and will happen when digital services that we rely on close down?
- Survey with qualitative analysis



Dealing with Digital Service Closure

SANDY J.J. GOULD, Cardiff University, Wales, UK
SARAH WISEMAN, Viable Data, UK

People integrate digital services into their day-to-day lives, often with the assumption that they will always be available. What happens when these services close down? The introduction of services might be carefully planned, but their closure may not benefit from the same degree of consideration. A more developed understanding of the effects of closures might make it possible to minimize negative consequences for users. This paper builds on sustainability, digital memories, and collaborative-work research through an empirical investigation of service closure. Fifty-five participants completed a questionnaire that solicited experiences of service closure and attitudes toward prospective closure. Through a qualitative analysis of participant responses, we synthesized six themes that reflected the practical and emotional effects of service closure on people: disempowerment, disconnection, loss of capability, trust, time and effort, and notice periods. We make suggestions for ways that service features related to these themes might be managed during closure, but also identify less tractable challenges: as part of this investigation, we introduce and develop the concept of *service patinas* to describe the important but entirely service-bound data that contextualize digital artefacts.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; Collaborative content creation; Social networking sites; **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: service closure, service design, service patinas, collaboration, obsolescence, sustainability, digital memories, digital consumption objects, sharing, social media, data rights

ACM Reference Format:
Sandy J.J. Gould and Sarah Wiseman. 2022. Dealing with Digital Service Closure. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 504 (November 2022), 25 pages. https://doi.org/10.1145/3555617

1 INTRODUCTION

People rely on digital services to make friends, to collect and curate memories, to work, to meet potential partners, to buy things, and to while away a few hours. For many people, these services are essential infrastructure for their lives. This paper is concerned with what happens when these services close. How do people adapt to the practical impacts of service closure? Beyond just the logistical challenges of losing a service, what are the emotional effects of closures on users? These are increasingly important questions as collaborative work and leisure activities are increasingly mediated through remotely hosted, remotely controlled services. There is little guidance for service designers about the challenges that service closure creates for users, meaning the experience of closure can be more difficult for users than it might otherwise be. If we could begin to answer these questions, we could start to develop patterns for service closure that would minimize disruption to users.

Authors' addresses: Sandy J.J. Gould, goulds@cardiff.ac.uk, School of Computer Science and Informatics, Cardiff University, Abacws, Senghennydd Road, Cardiff, CF24 4AG, Wales, UK; Sarah Wiseman, sarahemwiseman@gmail.com, Viable Data, Banbury Road, Oxford, OX2 7HT, UK.

2573-0142/2022/11-ART504 $15.00
https://doi.org/10.1145/3555617

Example: service closure

Constraints:

- Ethical; hypothesis testing questionable
- Resource; opportunity sample
- Logistic; ‘spare time’ research
- Epistemological; knowledge too limited to successfully operationalize, necessarily prospective/retrospective

Example: service closure

We sent this for review at CHI 2022. One of the reviewers wrote:

”The paper starts by stating that participants that had not experienced closure of digital services were asked to 'imagine’ how they would feel if a service was taken away. This is not a valid approach to capturing user's opinions about a subject. A more valid approach would have been to ask participants to remove an App or refrain from using it for any length of time and then to consider how this made them feel.”

Example: service closure

R2's alternative:

- 'Digital detox' not the same thing!
- 'More valid' in what way? What would the construct be and what is the phenomenon of interest?
- What's a 'valid' approach to collecting opinions?

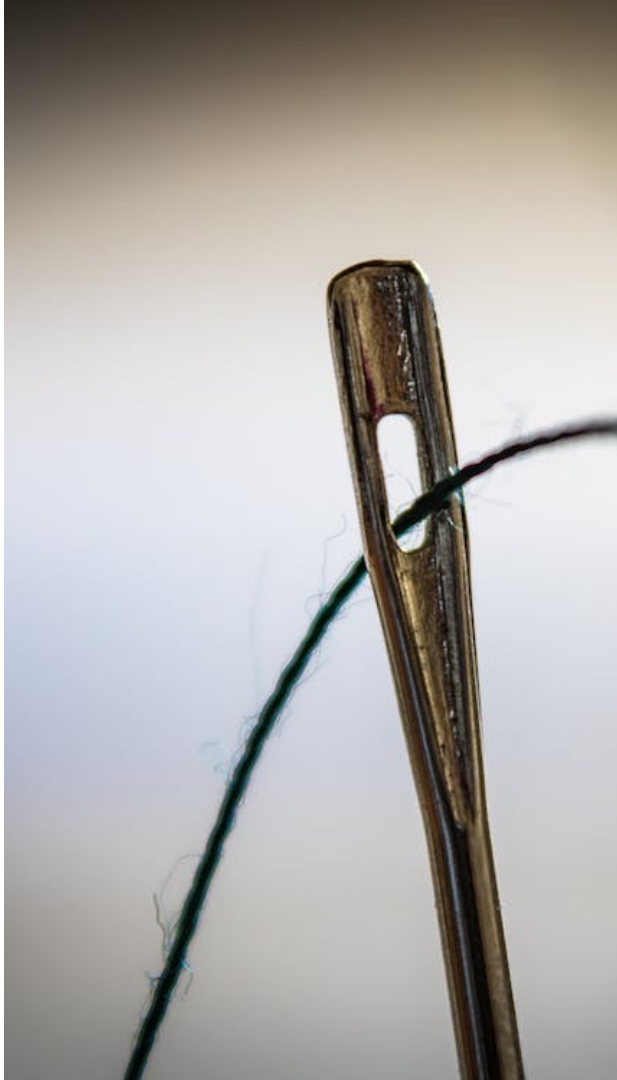


Photo: Francesco Ungaro Cortés on Pexels

Example: short links

Context:

- 2016 IJHCS paper
- Short-links are hard to type on smartphones. How could they be better designed?
- Monte Carlo modelling with an experiment to be run in a phone browser

Int. J. Human-Computer Studies 96 (2016) 38–53

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Short links and tiny keyboards: A systematic exploration of design trade-offs in link shortening services

Sandy J.J. Gould [a,*], Anna L. Cox [a], Duncan P. Brumby [a], Sarah Wiseman [b]

[a] UCL Interaction Centre, University College London, 66-72 Gower Street, WC1E 6BT, United Kingdom
[b] Department of Computing, Goldsmiths, University of London, London SE14 6NW, United Kingdom

ARTICLE INFO

Article history:
Received 7 March 2016
Received in revised form
23 July 2016
Accepted 25 July 2016
Available online 28 July 2016

Keywords:
Link shortening
Short links
Keyboard
Optimization
Typing
Data entry
Mobile
Touchscreen

ABSTRACT

Link-shortening services save space and make the manual entry of URLs less onerous. Short links are often included on printed materials so that people using mobile devices can quickly enter URLs. Although mobile transcription is a common use-case, link-shortening services generate output that is poorly suited to entry on mobile devices: links often contain numbers and capital letters that require time consuming mode switches on touch screen keyboards. With the aid of computational modeling, we identified problems with the output of a link-shortening service, *bit.ly*. Based on the results of this modeling, we hypothesized that longer links that are optimized for input on mobile keyboards would improve link entry speeds compared to shorter links that required keyboard mode switches. We conducted a human performance study that confirmed this hypothesis. Finally, we applied our method to a selection of different non-word mobile data-entry tasks. This work illustrates the need for service design to fit the constraints of the devices people use to consume services.

1. Introduction

Semantically meaningful URLs are often long. This makes them tedious to transcribe. QR codes obviate the need to type, but are not human-readable and have a number of issues with usability (Shin et al., 2012) and security (Vidas et al., 2011). Link shortening services like *bit.ly* provide a compromise: the process of typing complex and lengthy URLs is accelerated and human readability is preserved. Unfortunately, many shortening services exhibit little consideration for how links might be made quick and easy to type. The outputs of shortening services usually contain a mix of numbers and mixed-case letters. On space-constrained mobile devices, entering these characters requires changing the keyboard from lowercase mode to number mode or uppercase mode. Making these mode switches to access different characters is particularly time consuming and error prone (Greene et al., 2014). Shorter links, therefore, may not necessarily be faster to type.

Generating short links that do not require keyboard mode switches necessarily means using a smaller selection of characters. For a given link length this means fewer unique links can be generated. This reduction in the number of options for each character of a link can be mitigated by increasing the length of links. But how much longer would such a link need to be? In this paper we model the process of text entry on three widely-used mobile platforms. We use simulations to systematically explore the output of a popular shortening service, *bit.ly*. To see whether the predictions we make based on our model hold true in reality we test them in a game-like human performance study.

We found that link shorteners trade-off link length and entry difficulty. Most shorteners optimize too aggressively for link length; their output is awkward to transcribe. We show that links can be made easier to type with only a modest increase in their length. Given the increasingly limited functional utility of ultra-short links on services like Twitter, link shortening services should prioritize making links easier to type.

1.1. Related work

Efforts to make text entry easier for people have mostly focused on improving entry interfaces, like keyboards (e.g., Cheng et al., 2013; Leiva et al., 2015; Oulasvirta et al., 2013). Another way to improve text entry is to adapt input interfaces to accommodate the kinds of input that they are most likely to receive. Wiseman et al. (2013) showed that in hospitals the distribution of digits entered into devices is not random. Input interfaces that are designed for a particular set of possible inputs perform better than standard interfaces that do not take account of the strings that are likely to be entered (Wiseman et al., 2013). The preponderance of third-party

* Corresponding author.
E-mail address: s.gould@cs.ucl.ac.uk (S.J.J. Gould).

http://dx.doi.org/10.1016/j.ijhcs.2016.07.009
1071-5819/

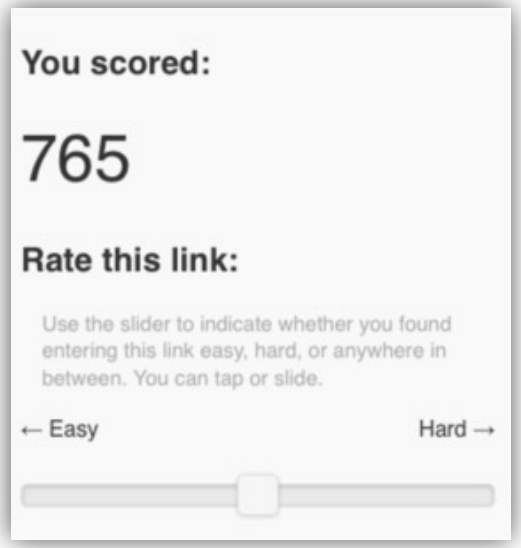
Example: short links

Constraints:

- Ethical; ⛔ intercepting keyboard input ⛔
- Resource; random side project – building custom keyboards too much.
- Logistic; gamification required synchronous group sessions
- Epistemological; deficient knowledge of optimisation strategies.

Example: short links

- After people had typed in a short link (e.g., `http://bit.ly/e4Rt5rr`), they were asked to rate how hard or easy they found entering it using a slider.
- A participant mentioned that they didn't think this could measure anything about subject experience, and I have been thinking about that since.



Example: short links

Possible alternative:

- Simply *do not attempt* to measure subject experience of difficulty!
- We had behavioural measures, why would an experiential ones be useful for our research questions?
- So why did I include it?

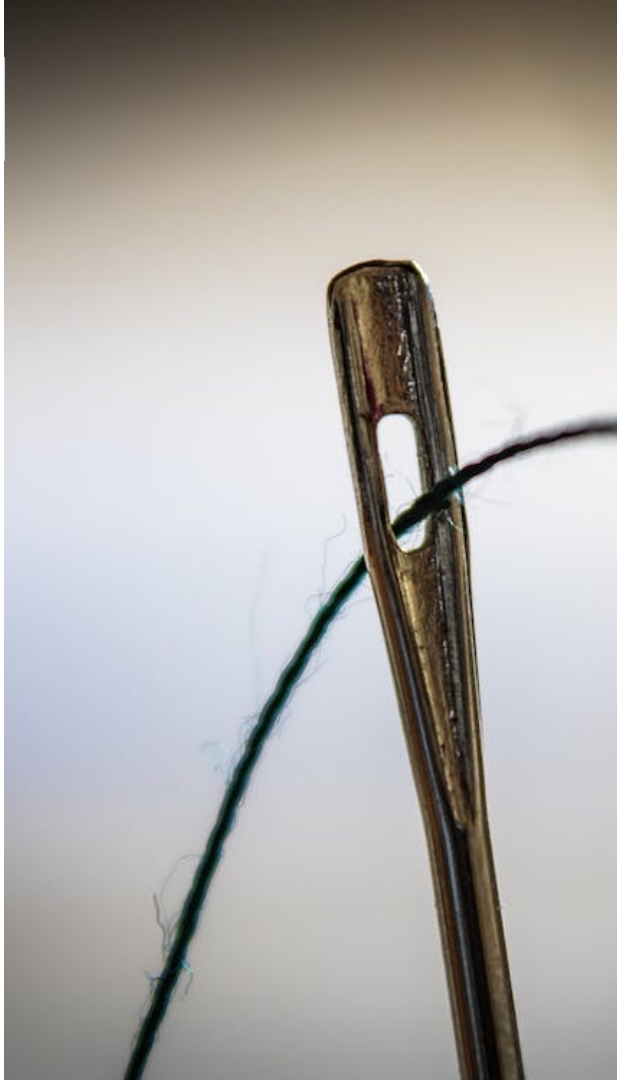


Photo: Humphrey Muleba on Pexels

How do we decide what to measure?

- Latour noted there was nothing special or magically scientific about laboratories. They are like other workplaces.
- The gap between what we *do* measure and what we *would like* to measure is subject to 'irrational' influences.
- I explored some of these influences in a CHI 2022 paper.

Consumption experiences in the research process

Sandy J.J. Gould
goulds@cardiff.ac.uk
School of Computer Science and Informatics
Cardiff University
Cardiff, Wales, UK

ABSTRACT

Data collection is often a laborious enterprise that forms part of the wider craft skill of doing research. In this essay, I try to understand whether parts of research processes in Human-Centred Computing (HCC) have been commodified, with a particular focus on data collection. If data collection has been commodified, do researchers act as producers or consumers in the process? And if researchers are consumers, has data collection become a consumption experience? If so, what are the implications of this? I explore these questions by considering the status of craft and consumption in the research process and by developing examples of consumption experiences. I note the benefits of commodity research artefacts, while highlighting the potentially deleterious effects consumption experiences could have on our ability to generate insights into the relations between people and technology. I finish the paper by relating consumption experiences to contemporary issues in HCC and lay out a programme of empirical work that would help answer some of the questions this paper raises.

CCS CONCEPTS

• **Human-centered computing**;

KEYWORDS

research methods; data; commodity data; data collection; commodification; consumption; craft; crowdsourcing; science and technology studies

ACM Reference Format:

Sandy J.J. Gould. 2022. Consumption experiences in the research process. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3491102.3502001

1 INTRODUCTION

This paper presents a novel critical perspective on the research process in human-centred computing (HCC)[1]. My thesis is that aspects of the research process have been commodified and that the process has, in some ways and at certain times, become a consumption experience for researchers. I think that this is something that should be scrutinised. The commodification of aspects of research may have benefits for researchers (e.g., payment handling, standard methods, access to participants), but the abstractions and subsumptions that come with commodification might risk encouraging practices that do not improve our capacity to generate new insight. This paper begins to develop conceptual prompts to help researchers think about how and why they choose particular research methods at different points of the research process.

I focus primarily on research in human-centred computing, but I also consider adjacent research domains where it makes sense to do so. The paper comprises five main sections. First, I attempt to provide some definitions for concepts like data, consumption, and commodity and I discuss the commoditisation of data and research in commercial and academic settings. This helps us frame a discussion of craft and consumption in research, before I move on to specific cases where I think parts of research processes have become consumption experiences. I conclude by asking *"Why does this matter?"*, relating the concerns of this paper to contemporary issues in HCI and proposing a programme of empirical work to answer some of the questions this paper raises.

2 DATA, COMMODITY, AND DATA AS COMMODITY

The goal of this paper is to make constructive criticism of research processes in human-centred computing and adjacent disciplines. To do so, I lean on concepts like 'data', 'commodity' and 'consumption'. These words mean very different things to different people. It is not my intention for this paper to provide foundational definitions of these concepts for human-centred computing research. Instead, I provide working definitions borrowed from other disciplines and adapted to the disciplinary context of HCC. These definitions help to focus the arguments that I make later in the paper.

2.1 Data

Data[2] has been studied from a number of perspectives in the human-centred computing literature. Some of this work has investigated how people track and make sense of data they collect about themselves [32, 111] or the data that third parties collect about them [125]. In parallel to these 'user-centred' investigations, other work has focused on the role of data in research (i.e., from a methodological perspective). This includes efforts to understand the reliability of research data [137], influences on the design of measurement [99],

[1] I use human-centred computing rather than human-computer interaction (HCI) in this paper because the topics I cover here apply to a variety of sub-disciplinary studies of relations between people and digital technology. HCI might be perceived by some researchers as only relating to the study of interaction. My intention is for this work to be applied more widely. For me, HCI research is a subset of HCC research.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3502001

[2] I use 'data' in the singular in this paper

‘Fast’ data

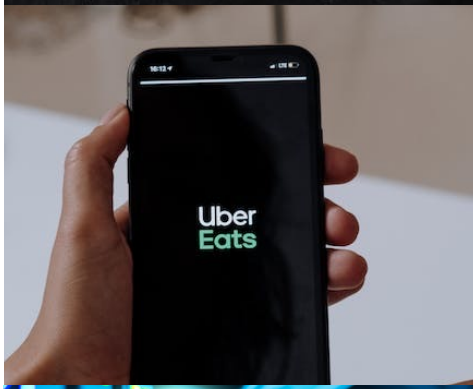
- Many crowdsourcing platforms offer research data quickly.
- How does this help to close the gap between what we can measure and what we want to measure?
- I argue it’s bound in a *consumption experience*.
- We’re susceptible to advertising!
- Other influences too – like publish-or-perish.

Speculations on ‘external’ influences

The paper focused on the immediate context of academic researchers.

What about the broader zeitgeist?

- ‘On-demand’ services
- Instant connectivity over the internet
- Commodification



Measurement out in the world

- Empiricism is a dominant mode of organisation ‘out in the world’.
- This is the case with work (e.g., the successors to Scientific Management), which is what a lot of my research focuses on.
- Unsurprisingly, developing empirical understandings of the workplace runs into the same challenges that we encounter in research, and has a few more to boot.

On Exactitude in Science

Jorge Luis Borges, *Collected Fictions*, translated by Andrew Hurley.

…In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography.

—Suarez Miranda,*Viajes de varones prudentes*, Libro IV,Cap. XLV, Lerida, 1658

Summary

- Measuring things is hard.
- We often have to use proxies.
- Some of these work better than others.
- They aren't always chosen because they work better.
- We should be reflexive about why we measure the things that we measure.

Thanks for your attention

http://sjjg.uk/uclic-23

goulds@cardiff.ac.uk

@sjjgo

Thanks to collaborators on the papers



Photo: Pixabay on Pexels