

Research on dark patterns: a case study of the Princeton CITP formula

Arvind Narayanan

Apr 27, 2021

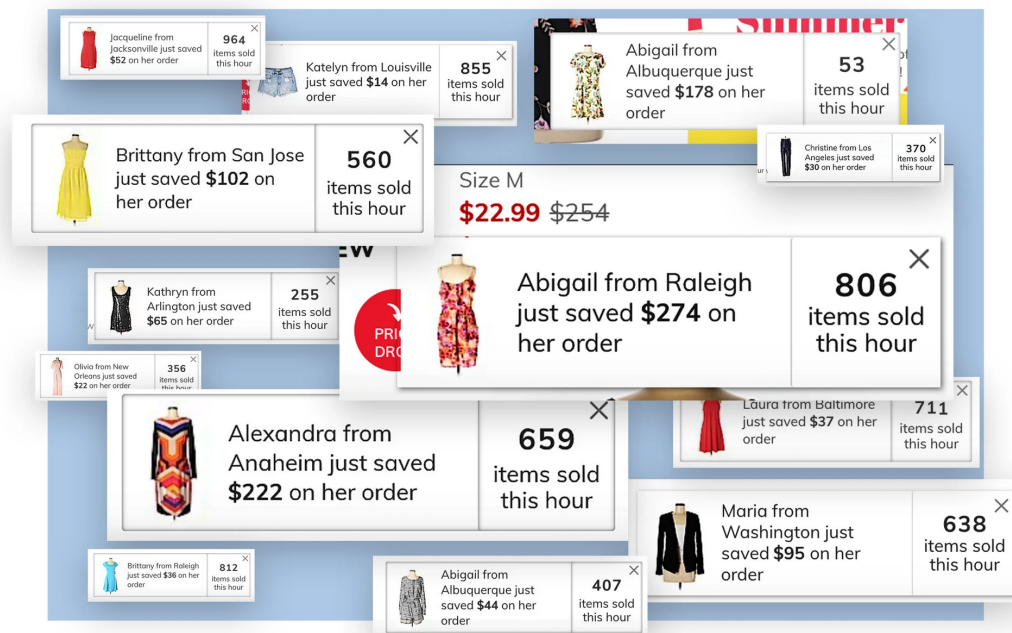
At Princeton's [Center for Information Technology Policy](#), our mission is to understand and improve the relationship between digital technology and society. We are an interdisciplinary group of about 40 researchers whose expertise is centered in computer science but includes the social sciences, law, and humanities. This case study describes our recent research on uncovering dark patterns — manipulative user interfaces — and our efforts to improve public understanding of the topic and advance policymaking. I use the case study to describe what is unique about CITP to those considering joining us and to serve as a blueprint for other universities considering establishing similar centers.

A brief history of dark patterns research at CITP

Dark patterns became a major area of research at CITP in summer 2018. Although the topic had been on our minds for a long time, several things happened within a span of a few weeks that convinced us to dive in. Senator Mark Warner highlighted dark patterns as a focus in his [list of policy proposals](#) for regulating tech companies, but the document only mentioned anecdotes to motivate the problem. Meanwhile we spoke with a prominent journalist who wanted to write about dark patterns but felt stymied by the absence of rigorous — especially large-scale — research on the topic.¹

This high level of interest convinced us that there was unmet demand for large-scale dark patterns research that would inform the public debate and shape tech policy. Through our previous work on [uncovering](#) online privacy violations, we had built up expertise on conducting large-scale investigative studies of websites. **Impactful research tends to happen when “supply” meets “demand”**, i.e. when the research team's expertise is well suited to fill a gap in existing knowledge.

¹ A [report](#) on dark patterns by the Norwegian Consumer Council came out during this time. While influential, it was limited to three companies' products.



The website ThredUp shows fake messages about purchases by other customers. The illustration is from a New York Times [article](#) about our work.

We decided to focus on shopping websites to keep things tractable. Within a few weeks, Arunesh Mathur, who led the project, had discovered most of the core ideas of the eventual paper: (1) many dark patterns tend to heavily rely on a few textual templates (e.g. “N people are looking at this right now”) and hence can be automatically detected; (2) these messages are often outright fake (for example, “N” in the above message might be a random number generated using JavaScript); (3) one reason dark patterns are so prevalent is that there are companies that offer dark patterns as a service (!).

Then we hit a wall.

Automatic detection of dark patterns turned out to be much harder than we had counted on. Our goal was to build a bot that could closely simulate a human shopper: identifying product pages on each website, visiting a selection of such pages, selecting product customization options, adding those products to shopping carts, and stopping just before checking out. Our analysis pipeline further had to automatically segment each webpage into its constituent user-interface elements and cluster and classify them so that we could extract dark patterns.

To their credit, Mathur and the other junior authors Günes Acar, Elena Lucherini, and Michael Friedman decided to hunker down to solve the automation issues and do as much manual analysis as necessary (the senior researchers on the project were Marshini Chetty, Jonathan Mayer, and me). In the end the project took over a year to complete. **Figuring whether to plow ahead or to quit, not knowing how far the finish line is, can be a scary decision for a research team:** one decision risks months or years of wasted work and the other decision risks missing out on a breakthrough.

We released our [study](#) in June 2019, over a year after we had begun exploring dark patterns. We found dark patterns on over 1,200 shopping websites, by far the largest exposé of dark patterns. We identified 22 companies that offered dark patterns as a service. We also contributed a new way to categorize dark patterns based on their latent attributes and map those attributes to cognitive biases.

The response to our paper was both swift and sustained. The press attention helped foster a debate on dark patterns among the public and among designers themselves. Unexpectedly, the paper was selected for a Privacy Papers for Policy Makers Award by the Future of Privacy Forum, an award usually given to law papers. We were invited to present the research at various senators' offices, the FTC, and the OECD. Over time, our paper has influenced a [study on algorithmic harms](#) by the UK's Competition and Markets authority, the [OECD's research on consumer policy](#), and the U.S. House Judiciary Committee's [investigation](#) on competition in digital markets.

The impact of our study caught us off guard, and we wondered why. In many areas of research including tech policy, **timing is almost everything**. It turned out that we had caught a cresting wave. Dark patterns were well known in the design community since at least 2010 (when the [darkpatterns.org](#) website was created). But dark patterns weren't nearly as widespread back then. If we'd done our study a few years earlier, there would likely have been far fewer findings to report. If we'd waited another year, we'd surely have been scooped.

Dark patterns exploded as a research topic in 2019 (in some small part because of our study). We realized there was a lot more to say. We no longer needed the large team that had assembled for the original investigation but instead worked in smaller groups based on expertise, interest, and bandwidth. We gained invaluable complementary expertise via legal expert Mihir Kshirsagar, formerly of the New York Attorney General's office, who joined CITP around this time, and sociologist Brandon Stewart, who began collaborating with us.

Our follow-up works are ongoing, and so far include a [paper](#) on what makes a dark pattern dark aimed at the research community, a [column](#) on the past, present, and future of dark patterns aimed at practitioners, and a contribution to a Stigler Center [report](#) on digital platforms aimed at policy makers. We've also engaged in numerous complementary activities: organizing academic workshops, engaging with regulators, giving talks, participating in podcasts, etc. **Once you acquire expertise on a topic, it's often helpful to share that expertise with different audiences in different formats.** This helps you maximize impact and avoid producing papers that lose focus by trying to do too much.

Finally, we are working on uncovering manipulation in related areas including [political emails](#) and online ads. Unlike the follow-ups described in the previous paragraph, these involve developing new areas of expertise rather than capitalizing on already acquired expertise. Like our original dark patterns paper, these are intensive empirical investigations that we expect will each require over a person year of work, and which we hope will pay off over a several-year timeframe.

The role of CITP and centers in general

The fact that this research took place at CITP was integral to our success every step of the way. Of course, research success is never guaranteed, and the history above makes clear that things could have turned out differently at many points. That said, institutional infrastructure can maximize the ability of research teams to take advantage of opportunities. Here we highlight a few ways in which centers in general, and tech policy centers in particular, can effectively provide such infrastructure.

Let me begin by clarifying what CITP isn't. There is no top-down orchestration of research. When I say that dark patterns research has been a major thrust, I don't mean that a committee decided so. It simply emerged through the choices of individual researchers (who sometimes self-organize into groups). CITP does the things I list below to align incentives among researchers and help them work better together, but having done so, it gets out of the way and trusts people to do impactful work. What CITP doesn't do is as important as what it does.

A nexus for disciplines

University departments tend to be organized by disciplinary expertise. In other words, departments are groups of people with similar training and expertise working on different problems.

But what's needed for effective research on tech policy (and many other topics) is for people with *different* disciplinary training to come together to work on the *same* problems. Centers can fulfill this need.²

Most of CITP's core members sit in the same building rather than being dispersed across different departments on campus. This is a key part of what enables interdisciplinary collaboration to happen organically. Many great ideas come about through spontaneous hallway conversations. Relatedly, a seemingly trivial but actually substantial barrier to collaboration is simply knowing what your colleagues' are excited about at any given moment and what their areas of expertise are. Regular interaction fostered by physical space helps fill this awareness gap.

The center also makes collaborations easier by creating a "trust boundary". A new collaboration, even within a university, is always a bit risky because academics are usually overcommitted and often drop projects or devote insufficient time to them — even if we don't like to admit it. But it's much harder to ghost someone if you work in the same building and see them daily or weekly.

Our dark patterns research drew from computer science (including computer security and human-computer interaction), law, and sociology. These and other disciplines including the humanities are represented at CITP.

Of course, CITP can't help much with the built-in headwinds for interdisciplinary research in academia. Even though our 2019 dark patterns paper was squarely within computer science, the fact that it melded perspectives from computer security and human-computer interaction made it tricky to get published. But over the years CITP researchers have learned to get better at navigating these structural issues.

Solving the Collingridge dilemma

A strong constraint on effective tech policy research comes from the so-called [Collingridge dilemma](#). If you study a technology too early in its lifecycle, you risk being off base because the social impacts are hard to predict (think of all the work on the dangers of nanotech and 3D printing). If you study it too late in its lifecycle, it becomes too entrenched and hard to control (arguably various applications of machine learning today).

² Of course, not all centers have to be organized this way. At Princeton, there is no statistics department, and expertise in statistical methods is dispersed among many departments. The [Center for Statistics and Machine Learning](#) serves as a nexus of this expertise.

A lot of the infrastructure that CITP provides can be seen as ways to mitigate these two problems. On the input side, we try to anticipate social impacts earlier by listening to outside experts; on the output side, we try to shorten the time lag between research and policy impact.

By doing both these things, we can try to maximize the Goldilocks zone for impactful tech policy research. Even with all these advantages, the zone might be just a couple of years; without them, the zone might be negative — it's often the case that by the time the research community at large starts to seriously tackle a tech policy problem, the window of maximal impact has already passed.

This is one reason why community is important in tech policy. Over the past 15 years, many, many CITP alumni have gone on to tech policy positions in government, nonprofits, industry, and other academic institutions. This helps us both learn about emerging topics and disseminate our findings more quickly through formal and informal channels.

Another helpful function of CITP in this context is funding. Raising project-specific external funding is generally too slow given that the window of opportunity to embark on a major project can sometimes be a matter of weeks. Indeed, we did not have external funding for our dark patterns work. CITP facilitates such projects by providing a funding backstop. CITP's funds come from a mix of philanthropic donations, corporate sponsorship, and alumni.

Finally, the [tech policy clinic](#) is another pathway by which CITP members generate impactful research and bring it to bear on the practical policy challenges faced by governments, companies, and nonprofits. The clinic is loosely modeled on clinics at law schools and aims to train the next generation of public-spirited technologists. In a typical activity, a "client" brings a problem to us, and students work with the client to develop potential solutions under the supervision of the clinic lead.

A bulwark against institutional bureaucracy

Universities like Princeton evolve slowly and have relatively static ways of structuring academic groups. Sometimes this structure is at odds with our vision of how to do tech policy research effectively. As mentioned earlier, interdisciplinary groups being physically colocated is a key aspect of our model. Further, we have a need for expertise in areas that have no school or department at Princeton, notably law. Having a center allows us to do things like recruiting legal experts and structuring ourselves differently, as well as collectively advocate for our mission

and our model. We are fortunate that the university as well as the engineering school and the school of public and international affairs understand and support CITP's way of doing things.

Here are two more key differences between CITP's structure and the university's more traditional structure. We recruit people in a number of nonstandard roles, including an associate director, a clinic lead, data scientists, and emerging scholars. These and other roles are critical to our success considering that much of our work doesn't take the form of publishing papers.

Second, we are agile in terms of our research topics compared to most academic scholars, which is unsurprising considering the Collingridge dilemma discussed above. This is traditionally difficult for faculty members to pull off since faculty evaluation tends to emphasize depth over breadth. But the CITP affiliation justifies adjusting the evaluation criteria for faculty hiring and promotion. Further, we have a relatively large cast of postdoctoral and visiting scholars, which allows CITP as a whole to be more agile than any individual researcher.

Despite these differences, we retain the strengths of the traditional academic model, notably the fact that most of our work is led by researchers working toward their PhD. Although agility is important, it is the depth of doctoral research that allows us to set ourselves apart from many of the other players in the tech policy space such as think tanks.

Concluding thoughts

CITP was founded in 2006 and owes its existence to the vision of Ed Felten who recognized the growing importance of tech policy 15 years ago. Since then we have gradually expanded and established a track record of impactful work in this area. But the importance of tech policy has massively outpaced our own growth. This document was written as we reflect on our history, plan our future, and observe the emergence of similar units at many other institutions. We hope you find our experiences useful for your own work in tech policy or related areas. If you're interested in joining us or collaborating with us, come to Princeton as a student, check out our [fellows program](#) and other [open positions](#), participate in our [events](#), or subscribe to our [mailing lists](#).