Another Subtle Pattern: Examining Demographic Biases in Dark Patterns and Deceptive Design Research

Katie Seaborn1*, Weichen Joe Chang1

1Tokyo Institute of Technology, Tokyo, Japan

Abstract
Dark patterns and deceptive designs (DPs) have captured the attention of scholars, designers, and practitioners the world over. An interdisciplinary and diverse group of people hailing from design, law, computer science, the behavioural sciences, and more have come together to unveil and critically respond to the problem of deception in user interfaces (UIs) of all kinds. Here, we examined another subtle, meta-level pattern well-known within the behavioural sciences generally: biases in who participates in DP research and knowledge production. We approach the topic of demographics from two angles: who is publishing research, and who is participating in it. Findings indicate demographic biases by author institution location and sampling regions, favouring English-speaking North America and Europe. We argue that we must address this hidden thread for the sake of inclusion and rigour in practice.

Keywords
Dark patterns, deceptive design, deceptive user interfaces, manipulative user interfaces, persuasive design, user interface design, scoping review

Figure 1: Continental demographic spread of authors by institutional location (left) and participants by continent sampled (right). Note that Russia is considered a part of Europe, count-wise. Representation is visualized as white (low) to purple (high).
1. Introduction and Background

Dark patterns and deceptive designs (DPs) [1, 2] refer to elements of the user interface (UI) that deceive, manipulate, and/or harm end-users for the sake of the purveyors. With over a decade’s worth of research and an acceleration in the past few years [1], the field of DP studies is starting to mature. Notably, global awareness has risen, perhaps due to leading reports such as those by the Organisation for Economic Co-operation and Development (OECD) [3] and legal action [4], as well as the general spread of DPs across a range of media.

DPs are fundamentally about people. We have the users—consumers, website visitors, online shoppers, gamers—who are subject to these designs. But we must consider diversity more broadly. Indeed, there could be another kind of subtle pattern within this much-needed body of work: bias in who is involved and where DP research is taking place. This is not unprecedented. Anglocentrism, or the treatment of English and Anglo cultures as the default, is a well-established bias across the sciences [5, 6]. HCI research [7, 8, 9, 10], including adjacent domains of study, such as human-robot interaction [11, 12], are also rife with “WEIRD” sampling biases: people recruited from nations broadly characterized as Western, educated, industrial, rich, and democratic. Findings from such WEIRD research are undistinguished and treated as the default for all of humanity, despite evidence to the contrary [13, 14]. We should not find it strange if the same is true for DP research. Nevertheless, this has not yet been examined.

We must also consider ourselves: the designers, authors, experts, practitioners, developers, engineers, researchers, professors, students, lawyers, legal experts, members of governmental and legal bodies, and everyone else who has been mobilized to take action on the problem. Sampling biases are linked to who make the decisions. Indeed, recent work at CHI has called for action and reflection on such matters of authorship inclusion and diversity [9, 15]. In short, we could be just as WEIRD as our participants [16]. We are also subject to Anglocentrism as participants in knowledge production. Here we are, writing in English and only English for our bedrock venue, the ACM Conference on Human Factors in Computing Systems (CHI). Such systemic Anglocentrism, to which we are beholden, could obscure the true diversity of knowledge creators. Finally, mapping out author diversity would reveal who is not (yet) participating and thus help us understand any sampling biases. As yet, no one has examined how DP research is faring.

The effects of unconscious biases, marginalization of perspectives, and lack of diversity in research is still being mapped out in HCI generally. The only work that exists within the DP space, to the best of my knowledge, is that of Hidaka et al. [17]. They approached the search for DPs in Japanese apps using Western taxonomies, but with an open mind. As a result, they discovered a new class of DP specific to the Japanese context. Another example is the term “dark pattern” itself. While Harry Brignull, who coined it [2], and CHI publisher ACM1 have made efforts to eliminate it, the term persists in the community, the public consciousness, and in law, where it has been codified and may be difficult to change. Yet, critical race scholars and anti-racism scientists have long called for an end to the use of racialized terms like “dark” and “black,” which are almost always associated with negativity and marginalization. CHI ’21 keynote speaker Ruha Benjamin has written at length about the ways in which racist ideas

1https://www.acm.org/diversity-inclusion/words-matter
are embedded in technologies—including our names for them [18]. In *The Disordered Chaos*, astrophysicist Chanda Prescod-Weinstein deconstructs the term “dark matter,” pointing out, as is the case for “dark patterns,” how the descriptor “dark” does not really describe the concept well, because it is not about colour or shade. In the case of “dark matter,” “absent” or “hidden” would make more sense. For “dark patterns,” “deception” or “manipulative” are more accurate (and recommended by the ACM). Perhaps “dark” means “hard to see” ... but, when it comes to “dark patterns,” it also means “bad,” thus *reifying the negative connotation with skin colour*, i.e., colourism, and *race*, i.e., racism. A final wrench: “dark” and “light” as modes of *visibility* are also ableist ways to frame deception in UIs: the assumption is that users are not blind or low-vision (B/LV). But this is not true [19]. Yet, the terms are already here, ubiquitous. Changing them now is an uphill battle. We must ask ourselves how we got here ... a situation that may be explained by Anglocentrism, WEIRDness, and other biases on our side of the fence.

In this position paper, we explore the degree to which this other subtle pattern exists in the research conducted on DPs so far: the *who* and *where* of DP research. We asked: **What are the demographics of researchers and participants in DP research?** We used the open data set produced from a scoping review on the latest DP research by Chang et al. [20], combined with an older data set by Gray et al. [1]. We found that demographic biases exist, both in terms of author institutional location and where participants have been sampled, notably from North America and Europe. We contribute (i) an initial empirical analysis of who is participating in DP research from two angles—researchers and participants—that indicates WEIRD and Anglocentric biases; and (ii) our open data set. We argue that we must acknowledge this subtle pattern and work to address it as a community of practice. This international, hybrid workshop, as a site of practice, is the ideal place to **gather a team of geographically- and linguistically-diverse researchers for this purpose**. Altogether, we aim to highlight the urgency of acting now by demonstrating yet another hidden pattern may be accruing in our field of study.

### 2. Methods

We conducted a basic content analysis [21] of the aforementioned open data set, creating a new version based on the screening phases of the review processes. The reconfigured data set is available here: https://bit.ly/anothersubtlepattern

#### 2.1. Data Set Generation

The original data set by Chang et al. [20] was constructed through two means: a scoping review based on the PRISMA Extension for Scoping Reviews (PRISMA-ScR) protocol [22] (covering work between September 13th, 2022 and November 30th, 2023) and integration of the earlier data set generated by Gray et al. [1], created through a systematic review process (covering work until September 13th, 2022). Both data sets drew from Google Scholar and the ACM Digital Library (ACM DL) using the search string “dark patterns” (in quotes). The original search returned 6,810 plus 183 results, and the second search returned 3,020 plus 158 results in Google Scholar and the ACM DL, respectively.

---

Items were included based on four criteria: written in English; containing the text “dark pattern”; published in an archival venue; and empirical research, at least in part. Items were excluded for several reasons: the wrong type, e.g., student thesis; not centred on DPs; inaccessible; and non-empirical reports. We refer to reader to Gray et al. [1] and Chang et al. [20] for the full details. Since the latest scoping review used the Gray et al. [1] review as a base, we started with that data set. However, we used the “screening phase” data, before exclusions based on the scoping review topic in Chang et al. [20] were made. We updated one item (a preprint now published). This left a total of 120 items.

### 2.2. Data Items and Extraction

Our goals were different from those of the original reviews, so we had to extract additional data. Since this is a work-in-progress, we focused on the demographics most readily available for extraction. Two researchers independently extracted data, with one leading (80%) and double-checking the extractions. Differences were discussed until consensus was reached. For participants, we extracted, where available: language of the study, study location, and participant nationality. For authors: author institution location. This data was not always available; we provide data on the gaps.

### 2.3. Data Analysis

Basic content analysis [21] was used to evaluate the demographics of authors and participants. This approach concerns data that is descriptive rather than interpretive, such as basic facts and term counts. In kind, the lead researcher extracted and counted labeled data about the demographics of authors and participants, wherever possible. Descriptive statistics, including counts and percentages, and visualizations of these data were generated.

### 3. Results

#### 3.1. Authors by Institutional Location

Across 120 papers, there were 367 unique authors in 438 author entries. We will count by author entry to best represent the weight of publishing among authors. 31 nations were represented based on institution location (Figure 1, left). These included: USA (152, 34.6%), Germany (50, 11.4%), UK (23, 5.2%), Japan (22, 5.0%), India (20, 4.6%), Norway (14, 3.2%), Denmark (14, 3.2%), Austria (14, 3.2%), Sweden (11, 2.5%), Netherlands (11, 2.5%), Luxembourg (11, 2.5%), Switzerland (10, 2.3%), France (10, 2.3%), Italy (9, 2.1%), Hong Kong (9, 2.1%), Australia (9, 2.1%), Ireland (8, 1.8%), Canada (8, 1.8%), New Zealand (5, 1.1%), Belgium (5, 1.1%), Slovenia (4, 0.9%), Singapore (4, 0.9%), Brazil (4, 0.9%), Thailand (3, 0.7%), Portugal (2, 0.5%), Lithuania (2, 0.5%), Turkey (1, 0.2%), Russia (1, 0.2%), Kazakhstan (1, 0.2%), Finland (1, 0.2%), and China (1, 0.2%).

68% (n=21) were Western nations. Author linguistic context (Figure 2a) was English (101, 84.2%), German (5, 4.2%) Japanese (3, 2.5%) Italian (2, 1.7%) Norwegian (2, 1.7%), and one each for Danish, Dutch, French, Scandinavian languages, Chinese, Indian language, and Lithuanian.
Figure 2: Descriptive statistics for author linguistic context, known and/or implied.

However, of these, 60 (50.4%) were implied based on the materials used (e.g., data sets, screen shots); 51 (42.9%) were stated in the paper (Figure 2a). Eight (6.7%) could not be determined.

Overall, there appears to be a pattern of Anglocentrism and WEIRDness in this corpus of work, notably English and USA author contexts.

3.2. Participant Samples

Half (n=60) of studies reported on a total of N=103,305 participant samples (Figure 1, right). Of these, the sample size of two (1.7%) were unreported. 42 (35%) were conducted online, and the context of 15 (12.5%) was unknown.

15 locations (65%) were Western regions. The location was unknown in 17 (14.2%) cases. Otherwise, representation by continent was: Europe (38, 31.7%), North America (22, 18.3%), Asia (15, 12.5%), and 2 or 1.7% for each of South America, Africa, and Australia. Specifically reported nations included USA (15, 22.7%), UK (9, 13.6%), Germany (5, 7.6%), Russia (4, 6.1%), India (3, 4.5%), China (3, 4.5%), Netherlands (3, 4.5%), Italy (2, 3.0%), Norway (2, 3.0%), Denmark (2, 3.0%), Canada (2, 3.0%), Ireland (2, 3.0%), Sweden (2, 3.0%), Hong Kong (2, 3.0%), Japan (2, 3.0%), France (1, 1.5%), Greece (1, 1.5%), Spain (1, 1.5%), Northern Ireland (1, 1.5%), South Africa (1, 1.5%), Mexico (1, 1.5%), Scotland (1, 1.5%), and Thailand (1, 1.5%).

As for authors, a WEIRD pattern appears to be present in the participant samples, with Europe and North America leading the way, specifically USA and the UK.

4. Discussion

Our initial analysis of the demographic spread in authors and participant samples indicates some diversity but otherwise clear patterns of Anglocentric and WEIRD biases. We must take these findings with a grain of salt. The selection criteria, which we will discuss as an agenda item, surely emphasized these patterns. At the same time, we found a similar level of WEIRDness across authors and participants as Linxen et al. [7] did for CHI: 73% to our 65%, plus 68% for authors. We also found a much greater degree of English as the linguistic context—84.2%—compared to HRI, which was 30.6% [11]. These differences may be due to other
factors: such as the relative maturity of each field, and thus the relative sample size of papers; differences between robots and DPs; the “EIRD” pattern in HRI, with great representation in Asia; gaps and varying conventions in reporting structures; and more. The exact numbers aside, we can recognize that DP research has some form and level of demographic biases.

We cannot provide evidence of any effects related to these patterns at this time. Future systematic review work may follow in the steps of Henrich et al. [14] by comparing frameworks and results by culture and region [23]. There may be more types of DPs tied to language and locale, ones that may not be found in Western or English contexts, as previously identified for Japan [17]. Opportunities to explore the relative presence or absence of culturally-sensitive DP varieties are on the raise. For instance, the Government of India, through the Department of Consumer Affairs, has put forward a notification under the 2019 Consumer Protection Act to prevent and regulate DPs 3. Perspectives on what constitutes a DP may also vary in ways tied to cultural values [24], social mores, and local markets. Understanding where, for whom, and how may aid in global adoption—or rejection—of design guidelines, legal statues, and public understanding of ethical practice. For example, initiatives on creating a shared lexicon and ontology of DPs have emerged as a hot topic for the research community, with implications beyond research, such as for regulatory sanctions [1, 25]. This is a necessary step with many benefits for various communities—design, academic, legal. However, we need to ensure that all voices have a say in defining lexical norms and shared vocabularies. Otherwise, we risk reifying Anglocentric and/or Western values and perspectives as representative of all humankind. As Henrich et al. [14] discovered, we cannot always paint humanity with such broad strokes.

We must also avoid problems, like the use of “dark,” that may trace back to the larger issue of not all voices being in the room—or being heard, even if there. Once ideas are formalized into structures, such as terms, and codified into academic databases, laws and statutes, and the public consciousness, they become difficult to shift [26, 27]. An anecdote: after carefully explaining the problem, the first author was told by a powerful media entity that they were going to move forward with the term “dark pattern” regardless because “it’s catchy.” They were apologetic but steadfast. What the first author fervently hopes is that the DP community will react differently. Let us take a moment to pause and reflect on how power operates here, within the global field of research and practice, particularly to whose benefit ... and whose detriment. As sociolinguists Jones and Singh [27] explain, we can “deliberately” and even radically (re)construct language, if we have the power and will to do so. What the present work reveals is that who has the power and who “we” are may not be an even spread.

For now, we offer ideas for next steps, both with respect to this initial work, and how we might better understand and tackle these biases as a community going forward.

4.1. A Meta Agenda

- **Finer-grained statistics:** Determining the WEIRDness of authors and participants is not always straightforward, and this is a work-in-progress. We will need to examine the “EIRD” factors and also author nationality, study location, and participant nationality. This may involve contacting authors for their demographics as well as further details on the participants, if known but not reported.

---

3[https://consumeraffairs.nic.in/theconsumerprotection/guidelines-prevention-and-regulation-dark-patterns-2023]
• **Finer-grained reporting**: Many details about the linguistic context of the research and where participants were sampled from was missing or implied. This represents a larger pattern of reporting issues across HCI and adjacent fields that we should seek, as a community, to correct [11, 7].

• **Extend collaboration efforts globally**: Established authors can reach out to researchers who may be interested in conducting research on DPs in their linguistic and/or cultural context. Workshops and events can target certain continents and involve people and entities from underrepresented regions known to have a stake in DPs or similar issues, such as General Data Protection Regulation (GDPR) in Europe.

• **Sample more broadly**: We can move away from replicating the status quo in HCI and the behavioural sciences more generally by sampling underrepresented regions and populations. Platforms such as SurveyMonkey and Prolific offer virtual access to many nations not yet represented in DP research, as far as we know. Language may be more tricky, but new collaborations with people who speak different languages and cutting-edge language tools, like ChatGPT, are sure to increase the reach of recruiting activities.

• **Include non-English publications**: A major limitation of the systematic reviews that led to the creation of the data set used in this study was the exclusion of non-English publications. For context, Gray et al. [1] excluded 25 and Chang et al. [20] excluded 39, although it is not known whether these texts would meet the other criteria for inclusion, notably whether they are about DPs. The next phase of this work—and arguably all systematic reviews on DPs—should address this limitation. This could involve new collaborations between researchers hailing from a variety of linguistic contexts, as well as the use of tools, such as ChatGPT.

4.2. Limitations

We acknowledge that our findings are limited in several ways.

The limitations of the original procedures used to create the data sets remain. Two major issues stand out. The first is the use of the keyword “dark patterns” in the title, abstract, and/or full-text. This may have excluded work characterized in other ways, such as “deceptive design” and “manipulative interfaces,” although many in our community use “dark patterns” and include alternatives in this meta data. A second issue is the exclusion of non-English papers. This was for practical reasons. Ironically, because of Anglocentrism, most work is published in English. Nevertheless, we missed work from other linguistic contexts not presented in or translated to English. This must be rectified in future work. Indeed, **one of the goals of this paper is to raise the issue and gather interested parties with knowledge of languages other than English**. We hope to do so in the workshop.

The analysis was also limited by the data available for extraction. For example, authors may not have reported all demographics. Additionally, the institution and location of study does not necessarily reflect the identities and origins of authors. Specifically, many young researchers and notably students study and publish abroad. A triangulated approach, one that combines institutional location and author identity data—perhaps requiring direct contact with authors—will need to be carried out. Future surveys of DP researchers can illuminate the true demographic spread of authorship.
5. Conclusion

Demographic biases appear to exist in DP research. While not particularly “weird,” we must recognize that a large segment of the research community and general population is not yet involved, and should be. Our key takeaways are to raise awareness among the DP community and inspire efforts to correct these biases. We must act before gates are established and what we know about DPs from a small sample of humanity becomes “common” knowledge.

Acknowledgments

Our gratitude to Peter Pennefather for reviewing a draft of this paper. We thank Gray et al. [1] for providing the initial open data set. Katie Seaborn conscientiously dissents to in-person participation at CHI; read their positionality statement here: https://bit.ly/chi24statement

References


[16] M. Meadow, D. Spurrett, It’s not just the subjects—there are too many weird researchers, Behavioral and Brain Sciences 33 (2010) 104.


6. Online Resources

We provide our open data set here: