9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Biased Bytes: On the Validity of Estimating Food Consumption from Digital Traces

Keywords: biases, validity, social media, food, images

Extended Abstract

Introduction. Given that measuring food consumption at a population scale is a challenging task [1], researchers have begun to explore digital traces (e.g., from social media or from food-tracking applications) as potential proxies. However, it remains unclear to what extent digital traces reflect real food consumption. Found data overcomes several of the biases typical of traditional methods, but may introduce new biases that threaten validity in their own ways [4, 5]. Despite their potential, error-prone and unreliable data and methods may do more harm than good if handled without the required caution. As our community increasingly relies on large-scale digital data sources, methods that offer insights into the validity of new measures thus become increasingly necessary [3].

Methods. The present study aims to bridge this gap by quantifying the link between dietary behaviors as captured via social media (Twitter) vs. a food-tracking application (MyFoodRepo). We design and apply a novel crowdsourcing framework for estimating biases with respect to nutritional properties and appearance, and we perform a case study of food consumption in Switzerland. Controlling for location, period, and food types, we contrast an extensive set of tweeted food images with images of consumed and tracked food.

We adopt a pairwise paradigm, where we confront human raters with pairwise choices (e.g., "Which of these two dishes looks tastier?") and later infer latent scores from the pairwise preferences. Following the Bradley–Terry (BT) model [2], we assume that each image i has a latent tastiness score $s(i)$ and that the probability that a rater will prefer image a over image b [image b over image a] in a pairwise comparison is proportional to the score of a [score of b]:

$$\Pr(a \succ b) = \frac{s(a)}{s(a)+s(b)}. \tag{1}$$

Given this setup, maximum likelihood estimation is used in order to infer the latent scores that best explain the empirically observed pairwise preferences. Let each MyFoodRepo image $i \in \{1,...,N_M\}$ have a an estimated tastiness score $\hat{s}(i)$, and each Twitter image $j \in \{1,...,N_T\}$ have an estimated tastiness score $\hat{s}(j)$. The tastiness bias $b(T,M)$ between food consumption measured with Twitter and food consumption measured with MyFoodRepo can be expressed as the difference in the average estimated tastiness scores measured via the respective data sources, T and M:

$$b(T,M) = T - M = \frac{1}{N_T}\sum_{i=1}^{N_T}\hat{s}(i) - \frac{1}{N_M}\sum_{j=1}^{N_M}\hat{s}(j). \tag{2}$$

Results. Applying our crowdsourced framework for bias estimation, we first find that food type distributions among social media foods vs. among consumed and tracked foods diverge. Controlling for the discrepant food-type distributions by studying food types individually, we find that Twitter still provides a biased view of food consumption as measured via food tracking. Tweeted food is, on average across food types, perceived as more caloric, less healthy, less likely to have been consumed at home, and tastier (example in Fig. 1), compared to actually consumed and tracked food.

For example, on average across food types, a median-tasty Twitter dish is among the top 26% tastiest MyFoodRepo dishes, and a median-caloric Twitter dish is among the top 34% most caloric MyFoodRepo dishes. While social media traces can be a reasonable proxy of tracked consumption for certain foods types, we find that, overall, food shared on social media and consumed and tracked food significantly diverge from each other.

Next, we discuss the relationship between three distributions: all foods consumed by the general population, food consumption estimated via MyFoodRepo, and food consumption estimated via Twitter (Fig. 2). The fact that there is a divergence between food consumption measured via the two platforms—food tracking and social media—implies that at least one of the two is not a faithful representation of true food consumption in the general Swiss population. We argue that it is less likely that food tracking is the main source of bias, and we conclude that researchers should be attentive and try to establish evidence of validity before using digital traces as a proxy for the true food consumption in the general population.

Discussion and implications. Measuring biases in digital traces is the first step towards correcting them and drawing valid conclusions despite their presence. Through a case study of the Twitter and MyFoodRepo platforms in Switzerland, contrasting tweeted food images with consumed and tracked foods, we provide grounding and first insights by controlling for location, period, and food types. We argue that researchers should be attentive and aim to establish evidence of validity before using digital traces as a proxy for the true food consumption of a general population. We conclude by discussing the potential sources of these biases and their implications, outlining pitfalls and threats to validity, and proposing actionable ways for overcoming them. The methods and findings reported here can inform researchers in their efforts to leverage digital traces for various applications. Researching human behaviors beyond food, our crowdsourcing framework can be used to measure many types of biases, including, but not limited to, politics and activism or behaviors important for health and well-being.

References

[1] Ann Bowling. Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, 27(3), 2005.

[2] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4), 1952.

[3] David Lazer, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford. Meaningful measures of human society in the twenty-first century. *Nature*, 595(7866), 2021.

[4] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, 2, 2019.

[5] Claudia Wagner, Markus Strohmaier, Alexandra Olteanu, Emre Kıcıman, Noshir Contractor, and Tina Eliassi-Rad. Measuring algorithmically infused societies. *Nature*, 595(7866), 2021.

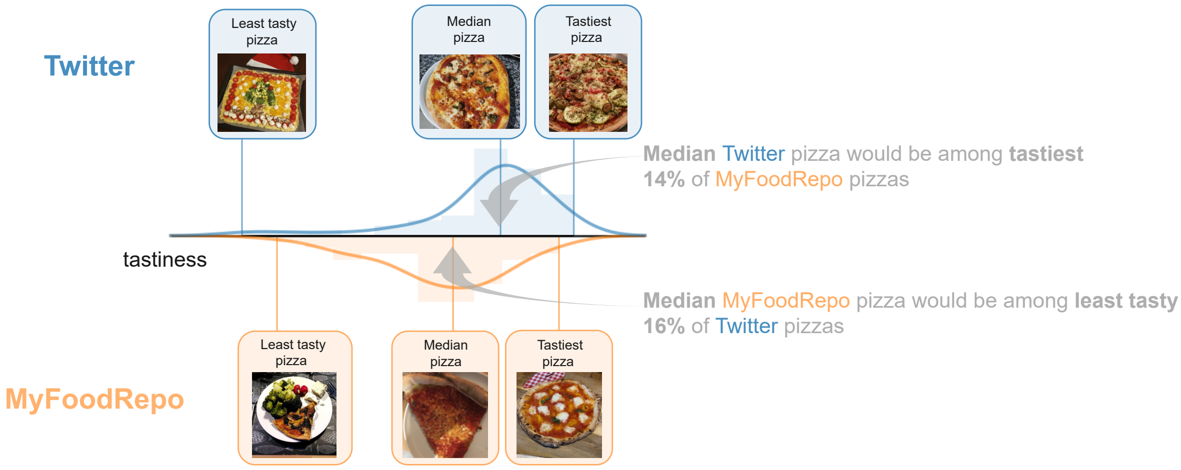

Figure 1: **Illustration of bias in perceived tastiness.** Perceived tastiness of tweeted food (*top*) vs. actually consumed and tracked food (*bottom*) of type "pizza". Histograms summarize tastiness scores estimated in our crowdsourcing framework. As illustrated, tweeted pizzas are perceived as considerably tastier than actually consumed and tracked pizzas.

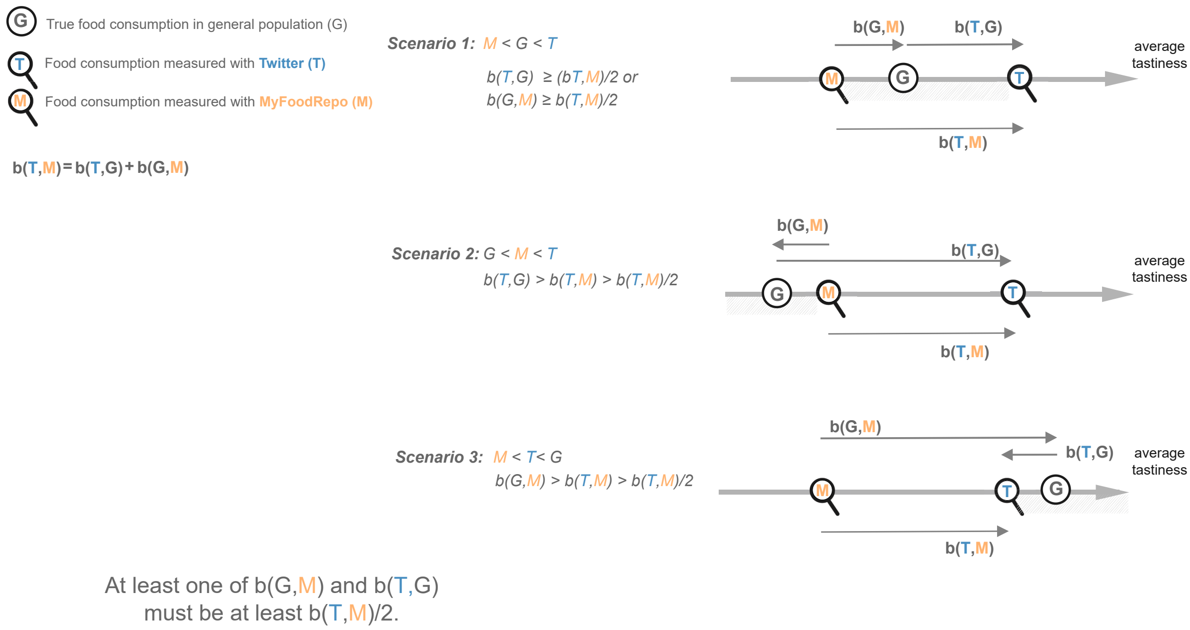

Figure 2: **Illustration of (tastiness) biases between true food consumption G in general population, tracked food M, and tweeted food T.** The bias between tweeted food and tracked food, $b(T,M)$, is characterized in our study, whereas $b(G,M)$ and $b(T,G)$ are unobserved. Although $b(T,G)$ and $b(G,M)$ are unobserved, at least one of $b(T,G)$ and $b(G,M)$ must be at least $b(T,M)/2$. The three illustrated possible scenarios depict situations where (1) $M < G < T$, (2) $G < M < T$, or (3) $M < T < G$.