Behav Ecol Sociobiol (2017) 71:24
DOI 10.1007/s00265-016-2254-4

ORIGINAL ARTICLE

Trustworthiness of online beer ratings as a source of social information

Petri T. Niemelä[1] · Niels J. Dingemanse[1]

Received: 30 June 2016 / Revised: 5 December 2016 / Accepted: 6 December 2016

Abstract
People increasingly use the internet as a source of social information. The pay-offs associated with using such information depend on its quality in terms of content and bias. A key question is therefore what information is contained in socially acquired information and whether it is biased. For example, social information may be biased due to conformism if people 'producing' (i.e. posting) information adjust it based on existing social information. We addressed these questions by focussing on ratings of beers posted by Finns on the internet, which people use as a source of social information when making consumer decisions. To model the information contained in beer ratings, we analysed a repeated measures longitudinal dataset of >130 000 beer ratings collected by 490 Finns and estimated key variance components. We decomposed variation in social information (i.e. ratings) into variation attributable to characteristics of the beer (beer identity, beer style, brewery and country of brewery), characteristics of the (individual) rater, variation caused by temporal effects and residual variation. Moreover, we compared blind with non-blind rating scores to evaluate whether conformism represented a source of bias. The majority (65.1%) of the variation in beer ratings was explained by beer characteristics, 9.5% by the identity of the rater and <0.5% by temporal effects; only 25.1% of the variance remained unexplained. Blind ratings were positively correlated with non-blind ratings, suggesting that conformism did not introduce a major bias. Our findings imply that beer ratings posted on the internet may represent a relatively unbiased and informative source of social information.

Significance statement
People use social information when taking behavioural decisions. Social information content may be biased due to conformism when people produce information non-independently. It is important to know whether social information is biased since social information that is of low quality is not useful. We quantified the information content of and bias in human beer ratings posted on the internet, which many people use as a source of social information. We show that beer ratings can be considered as an informative and unbiased source of social information: beer characteristics explain the majority of variation in beer rating scores, and blind and non-blind ratings were positively associated, implying that people do not produce biased ratings when scoring beers.

Keywords Human behaviour · Social information · Personal information · Conformism · Behavioural variation · Beer

Introduction

Humans and other animals base various behavioural decisions on social information, i.e. observed actions of other individuals (Danchin et al. 2004; Dall et al. 2005; Seppänen et al. 2007; Rieucau and Giraldeau 2011). Social and personal information usage has evolved because inherited genetic information is typically not sufficient in an ever-changing environment (Danchin et al. 2004). Whether individuals should use a potential source of social information should depend on the content of the social information: signals that do not contain reliable target information of interest should not be used

Communicated by M. Raymond

✉ Petri T. Niemelä
niemela@bio.lmu.de

[1] Behavioural Ecology, Department of Biology, Ludwig-Maximilians University of Munich, Planegg-Martinsried, Germany

Published online: 15 December 2016

(Giraldeau et al. 2002; Koops 2004; Seppänen et al. 2007; Rieucau and Giraldeau 2011). In general terms, information can be considered to be of high quality if all biases have small combined effects. In such cases, information realizes the intended potential of reducing environmental uncertainty (Koops 2004; Dall et al. 2005; Seppänen et al. 2007). For example, some animals like bumblebees use social information, and even prefer it over personal information, when the social cues are reliable, e.g. when social cues predict the presence of food with a high probability (Dunlap et al. 2016). Even though available information is rarely perfect, the quality and value of the social information for users will often be high when the information is produced by experienced or by high numbers of individuals, if the information is up-to-date and if the information user itself is naïve (Koops 2004; Laland 2004; Conradt and Roper 2005; Dall et al. 2005; Seppänen et al. 2007). All else being equal, the quality of the information is expected to improve its usefulness if it leads to more adapted (unbiased) decisions (Quinlan 1986; Keller and Staelin 1987; Dall et al. 2005; McLinn and Stephens 2006; Seppänen et al. 2007; Rieucau and Giraldeau 2011). In humans, decisions generally become more accurate when information is produced by a group of individuals (Keller and Staelin 1987; Dyer et al. 2008; Lorenz et al. 2011; Murr 2011). For example, people are able to collectively predict the outcomes of elections or estimate vaguely known facts with great accuracy (Lorenz et al. 2011; Murr 2011). Overall, information of low quality should not be used as such information can lead to maladaptive decisions with negative fitness consequences (Rieucau and Giraldeau 2011).

In our current human society, large amounts of ratings and questionnaire data are readily available on the internet and are increasingly used as a source of social information (Gosling et al. 2004; Flanagin and Metzger 2013; Metzger and Flanagin 2013). The main problem in interpreting online ratings is that it is difficult to discriminate between correct and incorrect information (Gosling et al. 2004; Flanagin and Metzger 2013; Metzger and Flanagin 2013). Information acquired from the internet is typically considered to be relatively unbiased given that it is generated by a large number of individuals and therefore reflects a collective consensus opinion (see above). However, conformism, which occurs when information producers adjust their judgements according to existing social information, can make the benefits of consensus opinions disappear by reducing the accuracy of group decisions (Lorenz et al. 2011). Additional features lowering the value of social information acquired from the internet is the uncertainty of what information exactly ratings represent, i.e. its information content.

Beer ratings are very suitable for studying the sources of variation in information posted on the internet. Hundreds of people rate thousands of beers online within and across seasons and years, where the same beer is often rated by multiple people and the same person often rates multiple beers. The resulting cross-classified nature of such datasets allows for the application of statistical models that partition the variance in ratings into underlying components (i.e. individual rater identity, beer identity, month and year identity and unexplained residual) and enables estimation of the magnitude of their respective contributions (Baayen et al. 2008; Bolker et al. 2009). If representing useful information, differences in ratings among beers should largely be attributable to beer-specific characteristics, e.g. beer style, brewery identity and country of brewery. Beer rating protocols are highly standardized and used globally to evaluate beers among professionals and beer enthusiasts producing millions of ratings to be used as a potential source of social information (Äyräväinen 2005; www.ratebeer.com; www.beeradvocate.com; www.olutopas.info). Nevertheless, we know of no scientific studies that have validated standardized beer ratings or quantified which aspects of beers explain variation in beer ratings based on longitudinal cross-classified datasets.

We conducted sets of data analyses to quantify the content of and bias in beer ratings. We first analysed a repeated measures dataset of >130 000 beer ratings collected by 490 individuals and estimated key variance components using cross-classified mixed-effects models. Ratings in this dataset were made with access to ratings made by other raters and could therefore suffer from conformism effects (detailed above). Moreover, ratings in this data set were not made blindly and could therefore be biased by expectations or reputation of the beer (Goldstein et al. 2008; Siegrist and Cousin 2009). We therefore, as a second step, also compared non-blind beer ratings to ratings collected blindly. The first set of analyses enabled us study in detail the sources of variation in beer ratings, which we achieved by decomposing the phenotypic variation in beer rating scores into components attributable to characteristics of the beer (beer identity, beer style, brewery and country of brewery), characteristics of the individual rater and residual variation. Since posted ratings were conducted over the timespan of many months and years, and because ‘mood’ is known to vary seasonally (Saarijarvi et al. 1999; Rastad et al. 2005; Kronfeld-Schor and Einat 2012), we considered that the value of the information (ratings) may change over time (Koops 2004; Dall et al. 2005). We therefore also considered month and year as potential factors explaining variation in rating scores. Our second analysis, where we compared blind beer rankings with non-blind rating scores, evaluated whether or not biasing effects caused by conformism, or the beer’s ‘reputation’, greatly biased internet beer ratings; we assumed that this would not be the case if blind and non-blind ratings of the same beer were positively correlated. We used a two-step approach to study the information content of beer ratings. First, we estimated how much of the variation was explained by beer identity, individual rater identity, month and year. If beer ratings represent an informative source of social information, we expected beer identity effects—not individual, month, year or residual—to explain the majority of the variation. Second, we aimed to understand what information was embedded in beer ratings by asking which well-known beer characteristics they predicted. We therefore expanded our original model by including beer style, brewery and country of the brewery. Doing so also enabled us to conduct a detailed examination of what information these ratings actually contained. This set of analyses thereby enabled us to estimate the best linear unbiased predictors (BLUPs) of rating scores for different beer styles and countries of brewery while controlling for biasing effects caused by differences between individual raters, months and years. We do not present estimates for specific beers or breweries because we do not aim to promote any commercial brands or products.

Methods

Data

We acquired beer rating data from a Finnish beer rating webpage (www.olutopas.info) on 13.04.2015 from the site authority. The beer rating data was completely anonymous; people rating the beers could not be recognized by anyone (including the authors of this paper). Altogether, our dataset contained 137,916 ratings from 6238 different unique beers of 77 different beer styles brewed by 1030 different breweries distributed over 53 countries and rated by 490 individuals between 1996 and 2015. Individuals rated on average (standard deviation) 281.5 (559.8) beers; each unique beer, style, brewery or country was rated on average 22.1 (27.6), 1791.1 (1879.3), 133.9 (346.6) and 2602.2 (6522.7) times, respectively. The raters varied in age (range 17–58 years); only 0.9% were women (this percentage is based on 222 individuals that chose to reveal this information). All people rating beers were Finns.

Behavioural data

We used the rating score as the behavioural target of our analyses. During a beer rating procedure, the individual goes through a highly standardized rating protocol, where aroma, appearance, taste, palate and overall verdict of the beer are estimated (in this specific order) using a numerical scoring system. Scores are given from 1 to 10 for aroma, 1 to 5 for appearance, 1 to 10 for taste, 1 to 5 for palate and 1 to 20 for the general impression of the beer. For 'aroma' and 'taste', beers that generally contain pure, more complex and long lasting notes (not necessarily all in one beer) obtain high ratings in these subcategories. Subcategory 'appearance' focuses subjectively on the colour and transparency of the beer and the structure of the foam of the beer. The 'palate' evaluates subjectively how the beer feels in the mouth focusing mainly on thickness, carbonation and stickiness of the beer. Beers that are overall in balance with all these features generally receive high rating scores, i.e. beers with too high or too flat carbonation or beers with a very thin body generally receive lower ratings. The 'general impression' of the beer represents the general verdict after evaluating all mentioned subcategories. The overall beer rating score represents the summed total of all subcategories, i.e. aroma, appearance, taste, palate and overall impression, and ranges between 5 and 50 points (Äyräväinen 2005; www.olutopas.info). Beers are rated on an absolute scale, measuring how much the person rating the beer overall enjoyed it (within all abovementioned subcategories), not relative to the beer's style (e.g. stout, pale ale, pale lager) (Äyräväinen 2005; www.olutopas.info; www.ratebeer.com). This makes all the rating scores comparable to each other irrespective of the classification of the beer (Äyräväinen 2005; www.olutopas.info; www.ratebeer.com). This kind of rating protocol is used globally as a standard for evaluating beers (Äyräväinen 2005; www.olutopas.info; www.ratebeer.com). All individuals rating beers have full access to the rating scores of others, as well as the collective rating scores for each beer before submitting their own rating scores. Beers are potentially scored under a large range of (physical and social) conditions, and this type of information was not available for our dataset. However, given that the same rater would perform ratings under a range of conditions, such an uncontrolled environmental variation would largely contribute to the residual variation in our statistical model.

Biases due to conformism or reputation

To investigate whether the variation in non-blind beer rating data is potentially driven by the reputation of the beer or by conformism among raters (versus true beer-specific characteristics), we compared a blind tasting data set (www.pastemagazine.com) with one of the largest non-blind, and collectively produced, beer rating databases available online (www.ratebeer.com). Those two datasets contained ratings for the same 50 Indian pale ales, which allowed us to compare them directly. The blind tasting data was produced by nine experts that did not know the identity of the beers and thus did not have access to any prior social information (www.pastemagazine.com). By contrast, for the non-blind data set, the beer raters had full access to collectively produced rating scores of the rated beers, and ratings of others, prior to submitting their own rating score. In the non-blind dataset, there were on average 829 (range 23–3650) ratings for each of these 50 Indian pale ales (www.ratebeer.com). Both abovementioned data sets came from the same country (USA). In both datasets, as well as in the main dataset used for this study, raters generally used the same beer tasting protocol described above with the exception that in the blind dataset, beers were ranked from 1 (best) to 50 (worst) instead of given a rating score. We do not have demographic data of the people that rated the beers for the two data sets described here.

Statistical analysis

We used cross-classified mixed-effects models to estimate how much of the variation in beer rating behaviour was attributable to individual, beer, beer style, brewery, country of brewery, month and year identity fitted as random effects. We first ran a model to estimate how much variation in rating score was explained by beer identity (Table 1) relative to individual identity, month, year and residual variance. We then also added beer style, brewery and country of brewery to determine how much of the beer identity effect was attributable to these beer-specific characteristics (Table 2). We controlled for variation in experience in rating behaviour in these models, where experience was defined as rating sequence within individual. Effects of experience occur because individuals learn to sense tastes or smells efficiently with repeated exposure, while experience may also change an individual's expectations (Parr et al. 2004; Siegrist and Cousin 2009; Dalenberg et al. 2014). In our data set, experience varied both within and among individuals because raters differed in their total amount of beer ratings (range 1–4982); preliminary analyses (following procedures detailed in van de Pol and Wright 2009) showed that within- and among-individual effects did not differ (results not shown). We therefore simply fitted (mean-centred) experience as a fixed effect covariate in our models. We calculated the percentage of variance not attributable to fixed effects that each variance component (individual, beer style, brewery, country, beer identity, year, month, residual) explained as a standardized metric (Nakagawa and Schielzeth 2010). Statistical significance of a focal fixed effect was based on the Wald F-statistic and numerator and denominator degrees of freedom. Statistical significance of a focal random effect was assessed by using a likelihood ratio test assuming an equal mixture of $\chi^2(df = 0)$ and $\chi^2(df = 1)$ distributions (Visscher 2006). This χ^2-distributed test statistic was calculated as twice the difference in Log Likelihood between the full model and a model where the focal random effect was removed (Meyer 1992; Wilson et al. 2010).

Mixed-effects models were fitted assuming a Gaussian error distribution. The response variable was expressed in standard deviation units prior to statistical analysis; all models were fitted using the statistical software ASReml 3.0.5. (Gilmour et al. 2009).

Results

Beer, individual, month and year identity all explained significant variation in beer ratings (Table 1). Beer identity alone explained the majority (65.1%), individual identity a modest amount (9.5%), while month (<0.1%) and year (0.3%) explained very little variation (Table 1). The unexplained variation remained modest (25.1%). Rating score was also affected by experience: people gave lower scores the more beers they had rated (Table 1). The relatively important contribution of beer identity in explaining variation among beer ratings implied that beer ratings were relatively informative within our data set.

The variance explained by beer identity dropped from 65.1% (Table 1) to 12.7% (Table 2) when beer-specific characteristics (style, brewery, country) were added to the model, representing the amount of unexplained residual variance among unique beers in the model. These beer characteristics explained 20.4% (style), 17.1% (brewery) and 11.5% (country) of the total variance in rating score, translating in 33.1% (style), 27.7% (brewery) and 18.7% (country) of the variance previously attributed to beer identity (Table 2). At the

Table 1 Sources of variation in beer rating scores derived from a univariate mixed-effects model where individual, beer, month and year identity were fitted as random effects and experience was fitted as a fixed effect covariate; we present fixed (β) and random (σ^2) parameters, F-statistics for fixed and χ^2 values for random parameters with their associated *P* values

Fixed effects	β (SE)	$F_{(\text{NUMdf, DENdf})}$	*P*	
Intercept	−0.022 (0.024)	$0.87_{1,65.6}$	0.356	
Experience	−0.025 (0.004)	$40.08_{1,961.9}$	<0.001	
Random effects	σ^2 (SE)	$\chi^2_{0.5}$	*P*	% (SE)[a]
Individual ID	0.087 (0.007)	14692.74	<0.001	9.5 (0.8)
Beer ID	0.593 (0.011)	124773.76	<0.001	65.1 (0.7)
Month ID	<0.001 (<0.001)	5.40	0.007	<0.1 (<0.1)
Year ID	0.003 (0.001)	75.60	<0.001	0.3 (0.1)
Residual	0.228 (0.001)	–	–	25.1 (0.4)

[a] Percentage of variance not attributable to fixed effects that was explained by the focal random effect

Table 2 Sources of variation in beer rating scores derived from a univariate mixed-effects model where individual, month and year identity fitted as random effects and where the beer identity effect (Table 1) was decomposed into key underlying characteristics (style, brewery, country of brewery); we present fixed (β) and random (σ^2) parameters, F-statistics for fixed and χ^2 values for random parameters with their associated *P* values

Fixed effects	β (SE)	$F_{(\mathrm{NUMdf,\ DENdf})}$	*P*		
Intercept[a]	−0.302 (0.077)	$15.47_{1,103.8}$	<0.001		
Experience	−0.026 (0.004)	$45.47_{1,1264.8}$	<0.001		
Random effects	σ^2 (SE)	$\chi^2_{0.5}$	*P*	% (SE)[b]	% (SE)[c]
Individual ID	0.087 (0.007)	14716.32	<0.001	10.5 (1.0)	–
Style ID	0.170 (0.029)	3257.6	<0.001	20.4 (2.9)	33.1 (4.3)
Brewery ID	0.142 (0.009)	1552.16	<0.001	17.1 (1.3)	27.7 (2.7)
Country ID	0.096 (0.032)	104.30	<0.001	11.5 (3.4)	18.7 (5.1)
Residual Beer ID	0.106 (0.003)	35421.36	<0.001	12.7 (0.7)	20.6 (1.8)
Month ID	<0.001 (<0.001)	6.62	0.004	<0.1 (<0.1)	–
Year ID	0.003 (0.002)	92.02	<0.001	0.4 (0.2)	–
Residual	0.228 (0.001)	–	–	27.4 (1.4)	–

[a] Analyses presented in Tables 1 and 2 are based on the same dataset; the estimated intercepts differ between Table 1 and 2 because the two models differ in random effect structure and because the levels associated with each random effect do not have the same level of replication

[b] Percentage of variance not attributable to fixed effects that was explained by the focal random effect

[c] Beer identity explained 65.1% of the variance in ratings (Table 1). We print here the percentage of this percentage that was attributable to characteristics of the beer (style, brewery, country) or remained unexplained ('unexplained beer ID' effects)

same time, beer ratings were also affected by other, unknown, beer-specific characteristics as a significant percentage (20.6%) of the among-beer variance remained unexplained (Table 2). The best linear unbiased predictors (BLUPs) that control for any bias due to individual, month or year effects are provided for beer style (Fig. 1) and country of brewery (Fig. 2).

Blind beer rankings (rank data: 1 = highest, 50 = lowest) predicted non-blind mean rating scores (adjusted to correspond to the scores in our main dataset: 5 = lowest, 50 = highest) (Kendall's tau = −0.229, *P* = 0.020), implying that reputation of a beer or conformism does not seem to cause overriding biases in beer rating datasets. In other words, beers that were blindly ranked highest were, on average, also given the highest collective rating scores in non-blind evaluations. We further note that the negative correlation between beer rankings and beer ratings should be interpreted as a positive association. This is because of the nature of the variables that are involved in this correlation: high rating scores (large numbers) in one predict high rankings (small numbers) in the other.

Discussion

The majority (65.1%) of the variation in beer rating behaviour was underpinned by beer identity effects, while individual, month and year explained little variation if any at all. The majority of the beer identity variance (79.4%) was attributable to the combined effects of beer style, brewery and country of brewery. Among-individual rater effects were small yet significant as some raters gave consistently higher rating scores compared to others. Nevertheless, the amount of individual variation was very modest (9.5%: Table 1) compared to expectations based on meta-analysis of repeatability of behaviour (37%; Bell et al. 2009). Importantly, blind beer rankings correlated with non-blind beer rating scores, suggesting that conformism or reputation did not strongly bias the beer ratings. This study thus implies that beer ratings made by Finns may largely contain meaningful information about beer characteristics. Because this social information thus seems unbiased, and because it can be acquired rapidly and with little cost, its usage should thus pay-off (Koops 2004; Dall et al. 2005; Seppänen et al. 2007; Rieucau and Giraldeau 2011).

Average estimates based on collective decisions made by multiple individuals, whether made by experts or not, are often considered to be unbiased (Conradt and Roper 2005; Dyer et al. 2008; Sumpter and Pratt 2009). Collective decisions made by groups of 100 or more individuals are almost never wrong (List 2004). Even groups consisting of fewer individuals make collective decisions that are much less prone to error compared to those taken by a single individual (Conradt and Roper 2005). Nevertheless, collective decisions can also be prone to error if the individuals producing information are not behaving independently. If information producers use social information themselves when taking decisions, consensus decisions may be biased away from the optimum (Lorenz et al. 2011). Such problems may readily occur when the judgments of others are visible to individuals producing

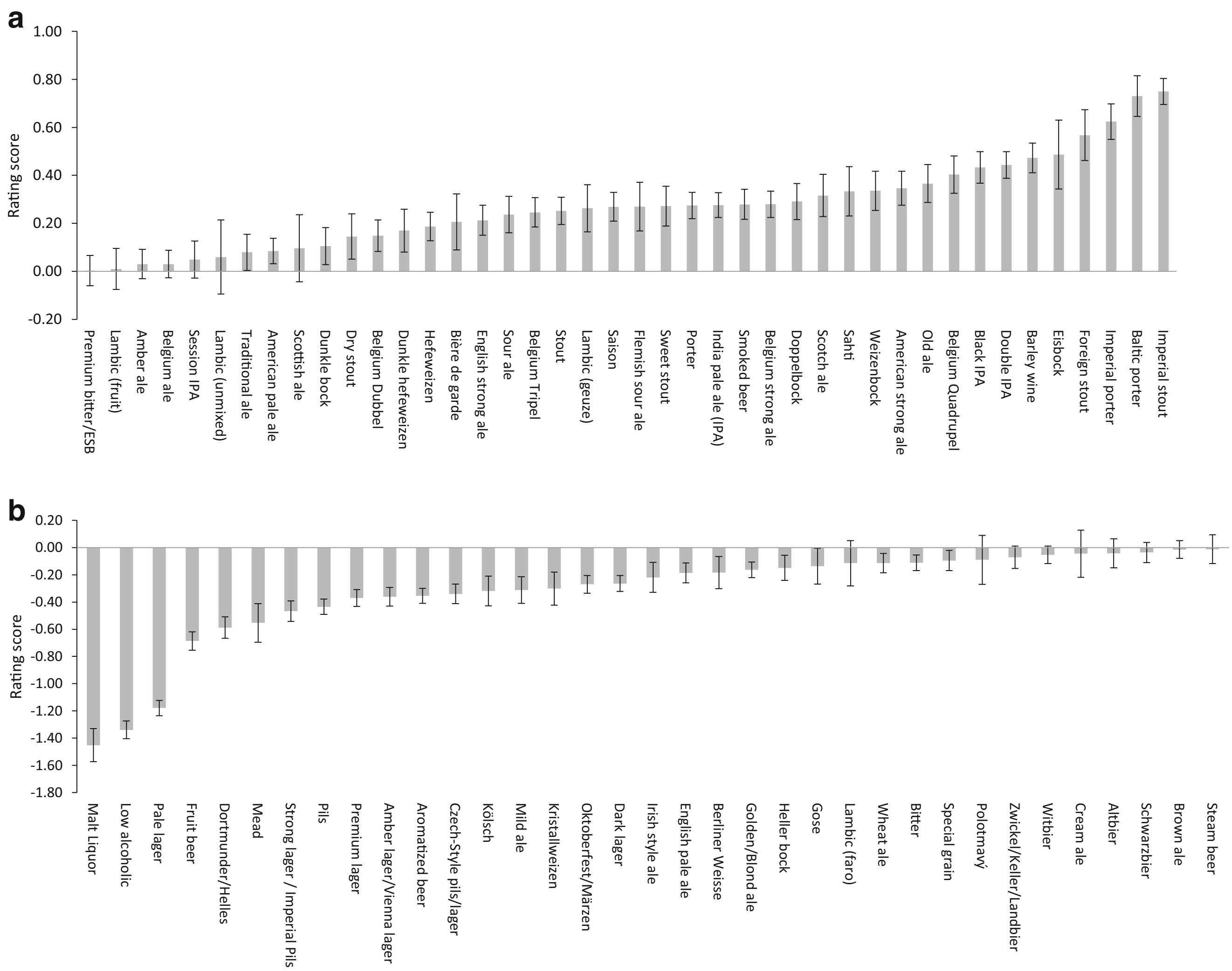

Fig. 1 Best linear unbiased predictors (BLUPs; mean plus standard error) of rating scores for various beer styles. BLUPs were extracted from the mixed-effects model detailed in Table 2 and thus control for biasing effects of individual, month, year, experience and other random effects. **a** Beer styles with above-average scores. **b** Beer styles with below-average scores

social information (Lorenz et al. 2011), which was the case for our beer rating data and may be the case for information posted on the internet in general. Nevertheless, the correlation between the blind beer rankings and non-blind beer ratings suggested that such conformism biases—if present at all—did not result in qualitative changes in relative appreciation. Interestingly, conformism would inherently lead to a relatively low amount of variance attributable to the individual identity of the rater. Since our analyses imply that conformism may not represent an important problem, the relatively minor individual identity effect present in our data may be caused by other processes. For example, individuals that rate beers may represent a non-random, self-selected, sample from the whole population (e.g. 99.1% was male) and thus show limited individual variation in behaviour compared to the population as a whole.

Since conformism did not bias the data at hand and the beer characteristics explained majority of the variation, the information harvesting strategy applied by people looking for information may largely define the amount of bias in this type of socially acquired information rather than the data itself. For example, if information users would acquire social information only from a single individual rather than from a collective, we would expect relatively biased and imprecise decisions as a result because individual (9.5%) and residual (25.1%) variation jointly explained a substantial amount of the variance. Across the animal kingdom, naïve individuals are more likely to rely on social information produced by more experienced conspecifics (Danchin et al. 2004; Laland 2004; Dall et al. 2005; Rieucau and Giraldeau 2011). This makes sense because copying experienced individuals can facilitate the usage of edible novel food sources or teach inexperienced individuals to avoid disliked, bad tasting or even harmful food items

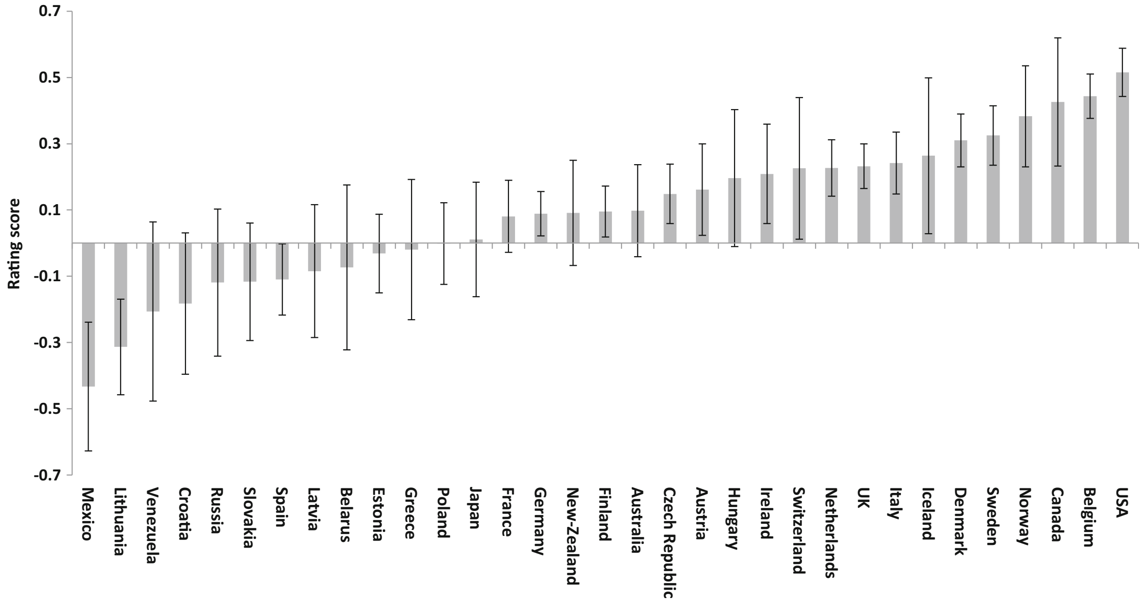

Fig. 2 Best linear unbiased predictors (BLUPs; mean plus standard error) of rating scores for the beer's country of brewery. BLUPs were extracted from the mixed-effects model detailed in Table 2 and thus control for biasing effects of individual, month, year, experience and other random effects. We print here BLUPs for 33 countries with ≥5 rated beers, though we note that our statistical analysis included all 53 countries available in the dataset

(Galef and Giraldeau 2001; Laland 2004). However, since the value of the information for the focal individual will partly depend on the ecological similarity between the information user and the information producer, information users should selectively favour the usage of information produced by conspecifics with a similar ecological niche (Seppänen et al. 2007). In humans, training leads to expertise that in turn affects one's tastes, pickiness and preferences (Bende and Nordin 1997; Parr et al. 2004; Goldstein et al. 2008; Dalenberg et al. 2014). Indeed, expectations are generally known to influence human judgment (Goldstein et al. 2008; Siegrist and Cousin 2009; Dalenberg et al. 2014). Our analyses also imply such a role for experience as people produced lower ratings the more beers they had rated previously. Experienced versus unexperienced people might, importantly, also differ in the types of beers that they prefer, implying that naïve individuals may benefit more from social information produced by relatively inexperienced raters. Moreover, the ratings presented in this paper were made by Finns, most of which were male. Though one could therefore argue that such a dataset is thus appropriately homogeneous, controlling for cultural differences, it would also mean that these collective ratings might be most beneficial for Finnish men interested in drinking a good beer. Indeed, cultural and gender differences in tastes and preferences for food generally exist for both genetic and learned reasons (Drewnowski 1997; Birch 1999; Garcia-Bailo et al. 2009; Werle et al. 2013), meaning that our results might not generalize across different cultures or genders. Along the same lines, while the authors—a Finn and a Dutchman living near Munich, Germany—fully agree with the low collective rating score of the "Oktoberfest" beer style (Fig. 1), one key question is whether our German colleagues would agree? It would therefore be useful to corroborate our findings by comparing more heterogeneous rating datasets, with a wider demographic distribution among raters, across countries or cultures to confirm general applicability.

Beer style (33.1%), country of origin (18.7%) and brewery (27.7%) explained the majority of the total phenotypic variation in rating score and 79.4% of the variation attributable to beer identity. This means that different beer styles (Fig. 1), breweries and countries (Fig. 2) attracted different average rating scores. In other words, those characteristics greatly shaped the content of social information. Beer styles and other characteristics differ in their taste, complexity and other features mainly because brewers use different amounts and different kinds of malts, hops and yeasts for different beers during their brewing processes (Zainasheff and Palmer 2007; Goldammer 2008), causing beers to differ in their taste and complexity. Moreover, methods used during the brewing process also differ between beers (Goldammer 2008). Breweries and countries, furthermore, have their own specific brewing traditions in terms of the types of ingredients or specifics of the brewing process (Jackson 1997; Goldammer 2008). Our findings imply that beer raters are able to distinguish between such, seemingly subtle, differences in beer characteristics and build a composite score that describes those differences in an unbiased way (Figs. 1 and 2). For example, styles that include heavily roasted malts such as stouts and porters, which are typically black or almost black in colour, and have chocolate, coffee and roasted tastes and aromas (Jackson 1997; Zainasheff and

Palmer 2007; Goldammer 2008), are rated highest by the Finns. Generally, types of lager like malted liquor, pale lager and helles, all of which include relatively small amounts of malts and hops and use brewing techniques and yeasts specific for those styles (Zainasheff and Palmer 2007; Goldammer 2008), were rated below average (Fig. 1).

In conclusion, this study demonstrated that online beer ratings made by Finns represent a relatively unbiased source of social information for other Finns: we did not find evidence for overriding effects of conformism or reputation. Indeed, beer characteristics explained the majority of the variation. More generally, individuals were able to evaluate and differentiate between beers with different qualities. Since the majority of the variance in ratings was attributable to beer characteristic rather than to (idiosyncratic) individual differences between raters or to unexplained residual variation, beer ratings should be informative for people interested in basing behavioural decisions on them. We should emphasize that the data analyzed here was collected by Finns and that most of the raters were men; this implies that our findings may not be generalizable across cultures or between women and men. Multicultural studies are thus clearly needed to further our understanding of the general applicability of our findings. Moreover, since the preferences of naïve and experienced individuals might not always match (Seppänen et al. 2007), we encourage readers to sample a large range of beers, from many styles and breweries, and from different countries to cross-validate whether social information on beer ratings provided by the Finns matches personal preference.

Acknowledgments PTN was supported by DFG (Deutsche Forschungsgemeinschaft, NI 1539/1-1). We thank Marika Pesonen and Anne Rutten for a stimulating dinner during which we tasted highly rated beers and conceived the idea for this study. We also thank two anonymous reviewers for the constructive comments on earlier versions of this article, Timo Kanniainen and Seppo Äyräväinen for checking beer 'facts' detailed in this paper and Seppo Äyräväinen for kindly providing us access to the beer rating data. We also thank Jim Vorel (from www.pastemagazine.com) and Joe Tucker (from www.ratebeer.com) for providing access for the data used in the analysis of conformism and reputation bias.

Compliance with ethical standards

Conflict of interests The authors declare that they have no competing interests.

References

Äyräväinen S (2005) Suomalaisen oma olutopas. Moreeni Ltd., Helsinki, Finland

Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang 59:390–412

Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. Anim Behav 77:771–783

Bende M, Nordin S (1997) Perceptual learning in olfaction: professional wine tasters versus controls. Physiol Behav 62:1065–1070

Birch LL (1999) Development of food preferences. Annu Rev Nutr 19: 41–62

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24:127–135

Conradt L, Roper TJ (2005) Consensus decision making in animals. Trends Ecol Evol 20:449–456

Dalenberg JR, Nanetti L, Renken RJ, de Wijk RA, ter Horst GJ (2014) Dealing with consumer differences in liking during repeated exposure to food; typical dynamics in rating behavior. PLoS ONE 9, e93350

Dall SRX, Giraldeau LA, Olsson O, McNamara JM, Stephens DW (2005) Information and its use by animals in evolutionary ecology. Trends Ecol Evol 20:187–193

Danchin É, Giraldeau L-A, Valone TJ, Wagner RH (2004) Public information: from nosy neighbors to cultural evolution. Science 305: 487–491

Drewnowski A (1997) Taste preferences and food intake. Annu Rev Nutr 17:237–253

Dunlap AS, Nielsen ME, Dornhaus A, Papaj DR (2016) Foraging bumble bees weigh the reliability of personal and social information. Curr Biol 26:1–5

Dyer JRG, Ioannou CC, Morrell LJ, Croft DP, Couzin ID, Waters DA, Krause J (2008) Consensus decision making in human crowds. Anim Behav 75:461–470

Flanagin AJ, Metzger MJ (2013) Trusting expert- versus user-generated ratings online: the role of information volume, valence, and consumer characteristics. Comput Hum Behav 29:1626–1634

Galef BG, Giraldeau L-A (2001) Social influences on foraging in vertebrates: causal mechanisms and adaptive functions. Anim Behav 61: 3–15

Garcia-Bailo B, Toguri C, Eny KM, El-sohemy A (2009) Genetic variation in taste and its influence on food selection. OMICS 13:69–80

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml user guide release 3.0., http://www.vsni.co.uk/resources/documentation/asreml-user-guide/

Giraldeau L-A, Valone TJ, Templeton JJ (2002) Potential disadvantages of using socially acquired information. Philos T Roy Soc B 357: 1559–1566

Goldammer T (2008) The brewer's handbook: the complete book to brewing beer. Apex Publishers, Clifton, Virginia, USA

Goldstein R, Almenberg J, Dreber A, Emerson JW, Herschkowitsch A, Katz J (2008) Do more expensive wines taste better? Evidence from a large sample of blind tastings. J Wine Econ 3:1–9

Gosling SD, Vazire S, Srivastava S, John OP (2004) Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. Am Psychol 59:93–104

Jackson M (1997) Michael Jackson's beer companion: the world's great beer styles, gastronomy, and traditions. Running Press, Philadelphia, USA

Keller KL, Staelin R (1987) Effects of quality and quantity of information on decision effectiveness. J Consum Res 14:200–213

Koops MA (2004) Reliability and the value of information. Anim Behav 67:103–111

Kronfeld-Schor N, Einat H (2012) Circadian rhythms and depression: human psychopathology and animal models. Neuropharmacology 62:101–114

Laland KN (2004) Social learning strategies. Anim Learn Behav 32:4–14

List C (2004) Democracy in animal groups: a political science perspective. Trends Ecol Evol 19:168–169

Lorenz J, Rauhut H, Schweitzer F, Helbing D (2011) How social influence can undermine the wisdom of crowd effect. P Natl Acad Sci USA 108:9020–9025

McLinn CM, Stephens DW (2006) What makes information valuable: signal reliability and environmental uncertainty. Anim Behav 71: 1119–1129

Metzger MJ, Flanagin AJ (2013) Credibility and trust of information in online environments: the use of cognitive heuristics. J Pragmat 59: 210–220

Meyer K (1992) Variance components due to direct and maternal effects for growth traits of Australian beef cattle. Livest Prod Sci 31:179–204

Murr AE (2011) "Wisdom of crowds"? A decentralised election forecasting model that uses citizens' local expectations. Elect Stud 30:771–783

Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol Rev 85:935–956

Parr WV, White KG, Heatherbell DA (2004) Exploring the nature of wine expertise: what underlies wine experts' olfactory recognition memory advantage? Food Qual Prefer 15:411–420

Quinlan JR (1986) Induction of decision trees. Mach Learn 1:81–106

Rastad C, Sjödén P-O, Ulfberg J (2005) High prevalence of self-reported winter depression in a Swedish county. Psychiat Clin Neuros 59: 666–675

Rieucau G, Giraldeau L-A (2011) Exploring the costs and benefits of social information use: an appraisal of current experimental evidence. Philos T Roy Soc B 366:949–957

Saarijarvi S, Lauerma H, Helenius H, Saarilehto S (1999) Seasonal affective disorders among rural Finns and Lapps. Acta Psychiat Scand 99:95–101

Seppänen JT, Forsman JT, Mönkkönen M, Thomson RL (2007) Social information use is a process across time, space, and ecology, reaching heterospecifics. Ecology 88:1622–1633

Siegrist M, Cousin ME (2009) Expectations influence sensory experience in a wine tasting. Appetite 52:762–765

Sumpter DJT, Pratt SC (2009) Quorum responses and consensus decision making. Philos T Roy Soc B 364:743–753

van de Pol M, Wright J (2009) A simple method for distinguishing within- versus between-subject effects using mixed models. Anim Behav 77:753–758

Visscher PM (2006) A note on the asymptotic distribution of likelihood ratio tests to test variance components. Twin Res Hum Genet 9:490–495

Werle COC, Trendel O, Ardito G (2013) Unhealthy food is not tastier for everybody: the "healthy=tasty" French intuition. Food Qual Prefer 28:116–121

Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEB, Nussey DH (2010) An ecologist' s guide to the animal model. J Anim Ecol 79:13–26

www.beeradvocate.com (accessed 7.12.2015)

www.olutopas.info (accessed 13.04.2015)

www.pastemagazine.com (accessed 7.12.2015)

www.ratebeer.com (accessed 7.12.2015)

Zainasheff J, Palmer J (2007) Brewing classic styles: 80 winning recipes anyone can brew. Brewers Publications, Boulder, Colorado, USA