A peer-reviewed version of this preprint was published in PeerJ on 4 January 2016.

View the peer-reviewed version (peerj.com/articles/1526), which is the preferred citable publication unless you specifically need to cite this preprint.

Woods AT, Michel C, Spence C. 2016. Odd versus even: a scientific study of the ‘rules’ of plating. PeerJ 4:e1526
https://doi.org/10.7717/peerj.1526

Odd versus even? An investigation on the impact of number of food items on plating preferences

Andy Woods, Charles Michel, Charles Spence

We report on the results of a series of large-scale computer-based preference tests (conducted at The Science Museum in London and online) that evaluated for the first time the widely-held belief in kitchens that food should be plated in odd rather than even numbers of elements on the plate in order to maximize the eye appeal of a dish. Participants were presented with pairs of plates of food showing odd versus even number of seared scallops (3 vs. 4), arranged in a line or as a polygon, on either a round or square white plate. No consistent evidence for a preference for 3 or 4 number of food items was found. The implications of these results are discussed.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1390v1 | CC-BY 4.0 Open Access | rec: 24 Sep 2015, publ: 24 Sep 2015

RUNNING HEAD: ASSESSING PREFERENCES ON THE PLATE: ODD VERSUS EVEN

Odd versus even?

An investigation on the impact of number of food items on plating preferences

Andy T. Woods, Charles Michel, & Charles Spence

Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, UK.

DATE: SEPTEMBER 2015

SUBMITTED TO: *PeerJ*

WORD COUNT: 11,109 WORDS

CORRESPONDENCE TO: Andy Woods, Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK; E-MAIL: andytwoods@gmail.com, TEL: +44 1865 271307; FAX: +44 1865 310447.

ABSTRACT

We report on the results of a series of large-scale computer-based preference tests (conducted at The Science Museum in London and online) that evaluated for the first time the widely-held belief in kitchens that food should be plated in odd rather than even numbers of elements on the plate in order to maximize the eye appeal of a dish. Participants were presented with pairs of plates of food showing odd versus even number of seared scallops (3 vs. 4), arranged in a line or as a polygon, on either a round or square white plate. No consistent evidence for a preference for 3 or 4 number of food items was found. The implications of these results are discussed.

KEYWORDS: PLATING; FOOD AESTHETICS; VISUAL PRESENTATION; EYE APPEAL; ODD VS. EVEN; GASTROPHYSICS; FOOD LIKING

ACKNOWLEDGEMENTS: We are grateful for suggestions for explanations for the results of Experiment 1 by Moritz Bernoully, Virtyt Gacaferi, Ellen Jarvis, Jens Karraß, Johanna Kuenzel, Daniel Lakens, Alejandro Salgado Montejo, Emily Snowden, Winni Theis, Catherine Transler, Pepijn Vemer, Kyra Woods, and Simon Woods. CS would like to acknowledge the AHRC Rethinking the Senses grant (AH/L007053/1).

INTRODUCTION

"Aller guter Dinge sind drei, nicht vier"

— *all good things come in threes not four* (popular German saying).

Disciplines that involve arranging items for aesthetic appeal share the common aim of seducing the eyes of the observer. From gardeners to chefs — plants and rocks at one end, through to food elements on the other — each discipline has its own insights concerning ways to enhance composition. These ideas are often transmitted orally, or sometimes, via books or guides. We believe that while the approach has historically been driven by experience, studying which of those guidelines are effective (and which translate cross-culturally) could benefit both these artisanal (or, may we say, artistic) disciplines, but also pose interesting scientific questions as to the nature of those biases.

One such guideline is the belief that it is better to present odd rather than even numbers of items (e.g. Van Tonder & Lyons, 2005)[1]. But is this anything more than 'an old wives' tale? Chefs often admit the importance of presenting odd numbers of elements on the plate, as recommended in chef's guides on the art of plating (e.g., Styler & Lazarus, 2006). However, to the best of our knowledge, this claim has never been put to the empirical test previously. So "Do odd vs. even numbers really matter when it comes to the visual appreciation of compositions?" And, to what extent can this difference influence the visual appreciation of the food, or maybe even the actual enjoyment of the food? One way of testing whether an odd number of items on a plate is preferred to an even number is to show participants two such plates of food, and ask which is preferred. Unfortunately, these two plates differ in more ways than just the number of items that they contain, which makes it hard to tease out the underlying driver of liking. We discuss such issues shortly.

[1] The idea of a preference for odd rather than even numbers of elements is not new. In the traditional art of Japanese rock gardens, for instance, ancient texts mention the importance of preferring odd vs. even numbers (Shimoyama, 1976; Van Tonder & Lyons, 2005), not only in terms of odd-numbered groups of rocks, but also that the different clusters of rocks should also be an odd-number.

In terms of food, there is very little research on the topic. Furthermore, none of this research has focused on the question of odd versus even numbers. One of the only researchers to tangentially do so, Bajaj, in his doctoral thesis, gave 215 participants the option of eating a piece of chicken cut into either 4 pieces, or left as a single piece (Chapter 3 Experiment 1, 2013). Although significantly more participants chose the 4-item dish over the 1-item dish than expected by chance (148, $p<.001$), no difference in pleasantness was reported between these individuals and those deciding on the 1-item dish. In a second study, 301 participants were randomly assigned to meal type (a bagel served in 4 pieces vs. whole) but pleasantness did not vary across the participant groups. The issue with Bajaj's study, in relation to 'odd versus even' number of items on the plate, is that the number of food items were quite different (1 vs. 4). We would expect, and will discuss next, a range of issues that could sway one's opinion on dish preference, which most likely are only exacerbated by large differences in the number of items / sub-portions.

For example, Geier, Rozin, and Doros (2006) put forward and demonstrated the concept of 'unit bias', where, when given the option to eat to satiation items of a small or large size, much smaller quantities of the small items were consumed than of the large. The consequence could be that, when asked to choose between plates of food, the most appetising portion will be that which matches one's current level of hunger (or dieting ambitions; see Forde, Almiron-Roig, & Brunstom, 2015, for a recent review on expected sensation in food selection). The logical consequence for preference between odd versus even number of items on a plate is that, if one portion appears bigger than the other, this will have a knock-on effect on choice selection.

However, even if portions are equated in terms of their calorific content, a variety of phenomena can act to influence just how large a portion of food may seem. For example, the size of the plate in relation to the food it contains has also been shown to influence perceived portion size thanks to the Delboeuf illusion (see McClain et al., 2014; Spence, Piqueras-Fiszman, Michel, & Deroy, 2014), where circles placed within a surrounding circle are thought of as larger than they actually are when there is a small size difference between the circles, but smaller than they actually are if the size difference in larger.

The visual balance of the composition can influence how we perceive and like food (for an overview, see Spence, et al., 2014, Michel, Velasco, Woods, & Spence, submitted) and presumably act to influence whether odd or even number of items on the plate are preferred. In terms of

balance, Zellner et al. (2011, p. 642) states that: *"The presentation of a plate of food can be thought of as 'balanced' if that plate of food looks like it would balance when placed on a narrow central pedestal. That is, the food is distributed in a manner around the central point such that the perceived heaviness in one area looks balanced by equal heaviness on the opposite side of the plate.*" Zellner et al. (2010) found that balance, in conjunction with food colour (or lack of it), influenced attractiveness of the visual presentation.

The artistic principles of visual harmony, including balance, contrast, emphasis, implied movement, pattern, proportion, rhythm, unity, and variety (Arnheim, 1988; Bouleau, 1980; Wilson & Chatterjee, 2005), could also influence food preference (Spence et al., 2014). Some aspect of harmony could be the driving factor in whether one prefers an odd versus an even number of items on the plate. Indeed, muddying the issue somewhat, the plate on which the food is presented could itself play in important role (as the 'frame' of the food).

Overview

We report on an experiment that is currently running at the Science Museum in London (see citizen science experiment). Participants were presented with photos of pairs of plates of food and asked to choose which of the plates they preferred. The pairs always consisted of individual dishes of food, one containing an even number of seared scallops and the other an odd number of the same food. We also assessed any interaction between odd/even, arrangement of the elements (line vs. polygon), and the shape of the plate on which the food elements happened to be presented.

The results of our first study revealed an intriguing interaction between odd/even and the shape of the plate on which the elements were arranged. There was no consistent evidence for our hypothesis that 3-items would be preferred to 4-items though. We explored these effects over a series of follow-up studies conducted online through Amazon's Mechanical Turk (MTurk). We controlled for the effects of crowding on the plate (Experiment 2), we equated portion size across the dishes (Experiments 3 and two further experiments reported as Supplementary materials), and we also tested for effects of portion size distortion (Experiment 4). The results of a Combined Analysis revealed that it was portion size that was the driving factor for both the participants at the Science Museum and those recruited via MTurk. These two groups of participants differed,

though, in terms of which dish (odd versus even) they preferred when we equated portion sizes over plates. Whilst the participants in the Science Museum study appeared to prefer 3-items at this 'equal portion-size' point, the MTurk participants preferred 4. In Experiment 5, we ruled out the possibility that this difference was not attributable to a small difference in the overall size of the two portions. We tested a third group of participants recruited through Prolific Academic in Experiment 6 to determine whether this group would have yet another equal portion size-point, but this was not the case (the values obtained from this group did not really differ from that of MTurk participants). We argue, though, that the ratio-effect most likely arises due to some difference in the characteristics of the populations tested.

EXPERIMENT 1

Here we tested the hypothesis that participants would prefer a dish of food containing 3-items of food over one containing 4-items.

MATERIALS AND METHODS

Participants

1816 individuals (1305 female and 509 male; 2 did not report whether they were male or female) took part in a citizen science experiment, conducted at the Science Museum in London during February to April 2015. The experiment could either be performed online (598 individuals)[2] or in an interactive digital platform at the 'Antenna Gallery', as part of an exhibition on the science of eating called 'Cravings'. The online participants were invited to access this experiment via the information page of 'Cravings' exhibition, and from the Science Museum's home webpage. At the museum's gallery, the digital platform was one of the attractions of the exhibition.

The median age of the participants was in the 16-34 years range (note that the participants specified if there age was <16, 16-34, 35-54, 55-74 or 75+; the respective counts in each group were 447, 880, 383, 92 and 12; 2 people did not report their age). All of the participants were informed about

[2] This experiment runs from the 20th of February 2015, until January 2016, see http://bit.ly/1MwGh35 to access the online experiment.

the nature of the study, and provided informed consent prior to taking part in the study and all of the studies reported thereafter. These studies have been approved by Oxford University's Medical Sciences Inter-Divisional Research Ethics Committee (approval # MSD-IDREC-C1-2015-004).

Stimuli

Scallops were chosen for the study, given that they are similar in shape (round) and size. Fresh scallops were seared in a hot pan with butter, in order to attain a light brown colouring. The same set of scallops was then placed and photographed on a white surface. Note that the scallops were photographed from a zenithal perspective, and with zenithal lighting, in order to avoid any shadow on the food. The scallop images were then cut and placed digitally on the different plates (square or round, photographed in the same way as the scallops). The stimuli used in this experiment are shown in Figure 1.

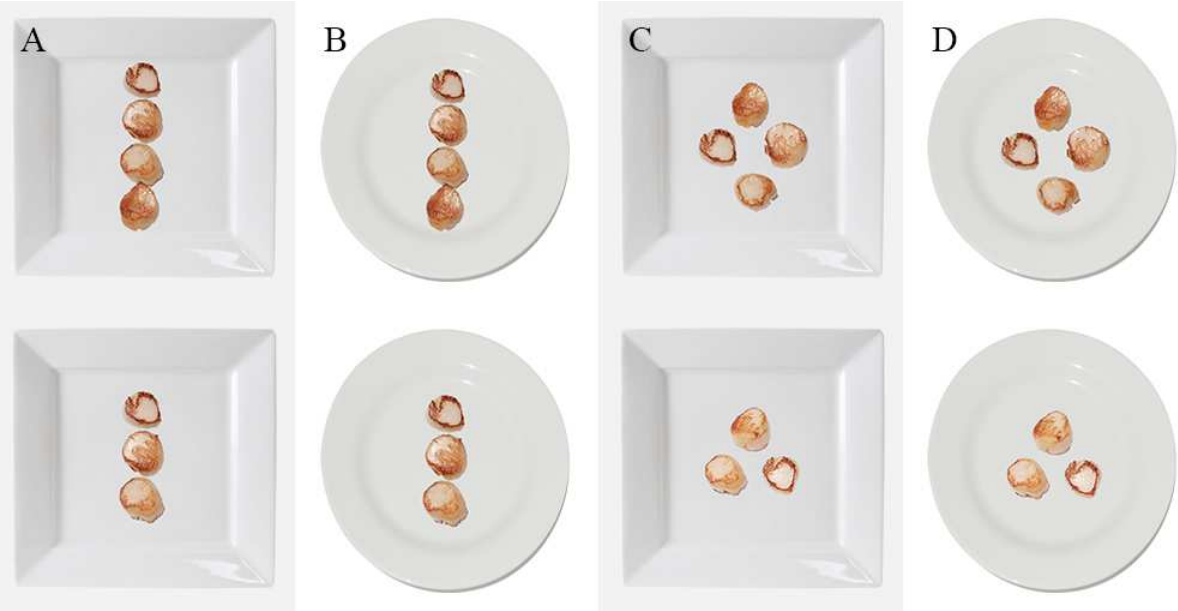


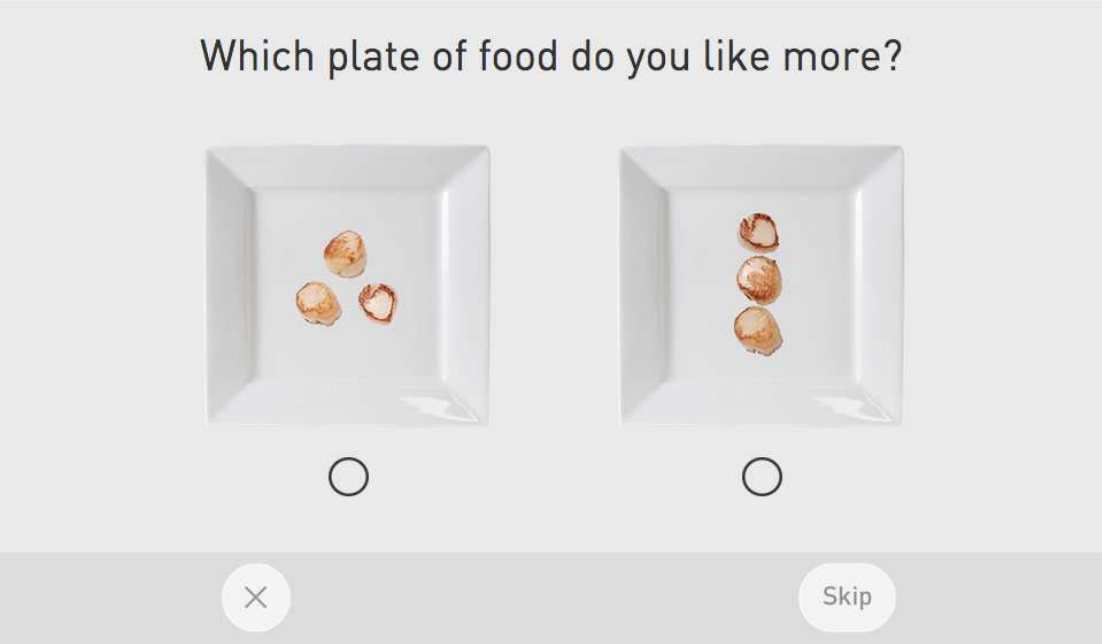
Figure 1. The 8 plates of scallops that were presented to the participants in Experiment 1. The plates were presented in pairs (specifically, the upper and lower image in each column was compared). The plates vary systematically in terms of the number of seared scallops (3 vs. 4), the arrangement of the scallops (line vs. polygon), and the shape of the plate (round vs. angular).

166 **Design**

167 The dependent variable was the preferred dish chosen by the participants.

168 **Procedure**

169 The participants who took part in this experiment undertook five or more different tasks, either
170 online, or at London's Science Museum 'Antenna Gallery'. The order in which the tasks were
171 presented and the different conditions was randomised, as were the left or rightward position of
172 the dishes. In the experiments reported here, 164 participants undertook two trials whilst the
173 remainder only completed one trial. The participants could either submit their answer by clicking
174 on a circular button placed right below the food image, leave the experiment by clicking on an 'X'
175 button, or go on to the next question by clicking on the 'Skip' button (see Figure 2).



176

177 Figure 2. The arrangement of elements shown to participants on a single trial.

178

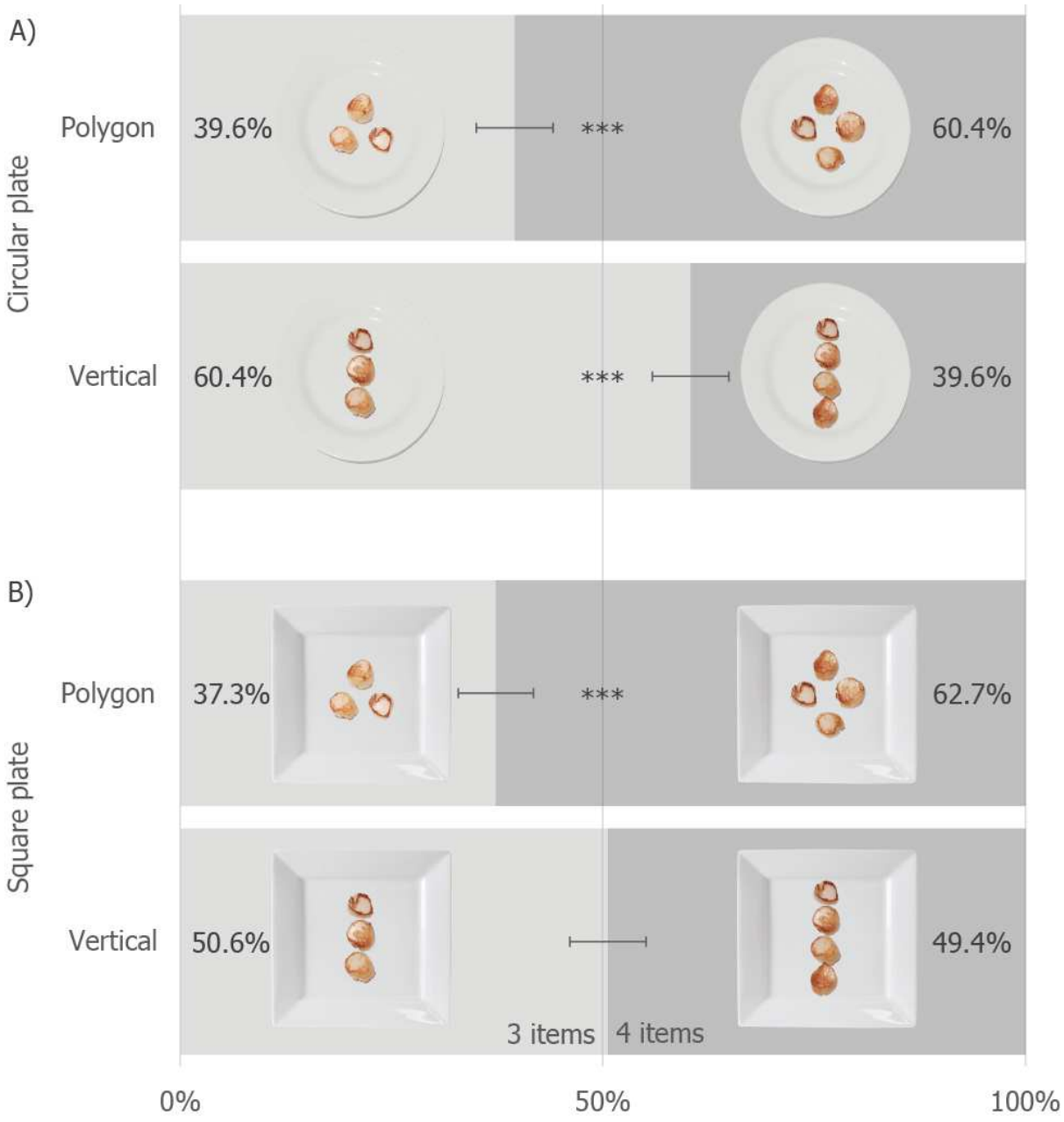
179 RESULTS

180 The results, split by condition, are shown in Figure 3. A log-linear analysis was performed, using
181 Plate Shape (circular, square) x food Arrangement (vertical, polygonal) x food Items (3, 4) as the

variables (the final model's likelihood ratio was $\chi^2(2)=3.27$, $p=.20$). The Arrangement x Items $\chi^2(1)=54.84$, $p<.001$, and Plate x Items interactions were retained by the model, $\chi^2(1)=6.63$, $p=.01$. Both interactions were explored by means of separate follow-up tests designed to assess whether 3 vs. 4 Items differed over the levels of the interacting factor.

In terms of the first interaction, in line with our hypothesis, 3 items that were arranged vertically were 1.24 times more likely to be chosen that 4 vertically arranged items ($p<.001$; with 428 picking the 4 item dish and 531 picking the 3 item dish; 95% CI 52.16%, 58.55%). Contrary to the hypothesis however, 4 items arranged as a polygon were 1.60 times more likely to be picked than 3 items arranged so ($p<.001$; with 578 picking the 4 item dish and 361 picking the 3 item dish; 95% CI 58.36%, 64.68%).

In terms of the Plate x Items interaction, 4-items on a Square plate were 1.24 times as likely to be chosen over 3-items on a square plate ($p<.001$; with 540 picking the 4 item dish and 426 picking the 3 item dish; 95% CI 52.70%, 59.06%). There was no preference when it came to the Circular plates (466 picking the 4 item dish and 466 picking the 3 item dish; 95% CI 46.74%, 53.26%).



196

197 Figure 3. The percentage of people preferring one dish over the over for each of the Plate and Arrangement
198 conditions (error bars are 95% CI, *** = p<.001). The light grey shading representing preference for the 3-
199 item dishes, and the dark grey the preference for the 3-item dishes.

200

201 DISCUSSION

The findings do not generally support the hypothesis that dishes with an odd number of items would be preferred over dishes with an even number of items. Although our analysis did find that 3-vertically orientated scallops were preferred over 4-vertically orientated scallops, inspection of Figure 3, shows that this effect only differed from that expected by chance when the scallops were plated on a circular plate. Thus, support for the hypothesis is actually more tenuous than that offered by the analysis. Indeed, overall, more evidence was found for 4-items being preferred over 3-items.

The lack of support for the hypotheses was unexpected, and after querying social media, several explanations were offered. One of the explanations that was proffered was that the portion sizes on 4-item plates were always seen as larger than those on 3-item plates. We tested for this in Experiments 3-6 by varying portion size by means of scaling the images of the scallops.

Two other issues were also suggested via social media. The first was that the four vertical items looked like substantially more food compared to those same number of items arranged as a polygon, and thus the dish was not preferred over the 3-item vertical dish as there was too much food on the plate. We go on to test this in Experiment 5 by asking participants how hungry they were, and testing whether this influenced the results. There was, however, no evidence for such an effect.

The second more subtle issue was that the 4-item vertical dish looked a little less elegant to us than the vertical 3-item dish, perhaps as the plate was seen as being too-full (some on social media even argued that the shape of the plate was distorted, becoming more oval). To test for this, in the next study, participants were exposed to dishes that were substantially larger than those used here, thus preventing the dishes from seeming too full.

EXPERIMENT 2: Testing for a crowded plate effect

Here we test the hypothesis that participants found the plate crowded for the vertically arranged dishes, which influenced how participants decided between a 4-item vs. a 3-item dish. To do this, we conducted a similar study with the same factors as the previous (number of items, food

alignment, and plate) and included an additional factor of plate size, albeit using a repeated measures design. Specifically, besides the 'regular' sized plate used in the previous study, we also collect data from trials where a much larger plate was used instead.

MATERIALS AND METHODS

Participants

One hundred participants[3] (35 female) were recruited from Amazon's Mechanical Turk to take part in the experiment in return for a payment of .40 US dollars. The participants ranged in age from 19 to 59 years (M = 32.0 years, SD = 8.4). The experiment was conducted on 6/06/2015, from 12:00 GMT onwards, and over a two-hour period. The participants took an average of 73 seconds (SD = 52) to complete the study. All of the participants provided their informed consent prior to taking part in the study.

Stimuli

The 5 unique scallop stimuli used in Experiment 1 were divided into separate transparent PNG files, as were the 2 plate stimuli. The 5 scallop images were individually resized so that they all contained approximately the same number of non-transparent pixels (the original number of pixels per scallop as 41193, 44817, 42869, 33272, 42199, with a standard deviation of 4449 pixels; after resizing the stimuli so that they approximately matched the average number of pixels per scallop, the pixels per scallop were 40828, 40791, 40904, 40796, 40642, with a standard deviation on 95 pixels). The scallop image dimensions were set to 50 x 50 pixels in the actual study. The plate images were set to 250 x 250 pixels.

The exact scallop images used in each dish were randomly determined, as were their set positions on the plate (care was taken so that the scallops were placed and spaced apart to resemble the vertical and polygonal arrangements that had been used in Experiment 1).

Design and Procedure

[3] A power analysis of the ratio of 3-item to 4-item preference for circular plated vertical/polygonal food from Experiment 1 (Generic Binomial Test, using G*Power 3.1.9.2) revealed that 90% power could be achieved in this study with an n of 62 or 64 (the former, vertically orientated food, the latter, polygonally-orientated food). We increased this to a sample size of 100.

The design was similar to that of Experiment 1 in that two plates of food were shown to participants on each trial, and the task was to decide on the plate that the participant most wanted to eat. Here, however participants undertook all 8 of the experimental trials, which differed in terms of the size of the plate shown (either both plates were large or regular sized), the shape of the plate (either both were square or circular), the arrangement of the food (either both were vertical or polygonal) and the number of food items (one plate there were 3-items, whilst there were 4-items on the other plate).

RESULTS

A log-linear analysis was performed, using Plate Size (regular, large) x Plate Shape (circular, square) x food Arrangement (vertical, polygonal) x food Items (3, 4) as the variables (the final model's likelihood ratio was $\chi^2(10)=3.54$, $p=.99$). Only the Arrangement x Items $\chi^2(1)=5.41$, $p=.021$ interaction was kept in the model. Separate Exact Binomial tests found that 4 items were preferred for vertically arranged items ($p<.001$; with 307 picking the 4-item dish and 93 picking the 3-item dish; 95% CI 19.20%, 27.70%) and for those arranged as a polygon ($p<.001$; 333, 13.22%, 20.78%).

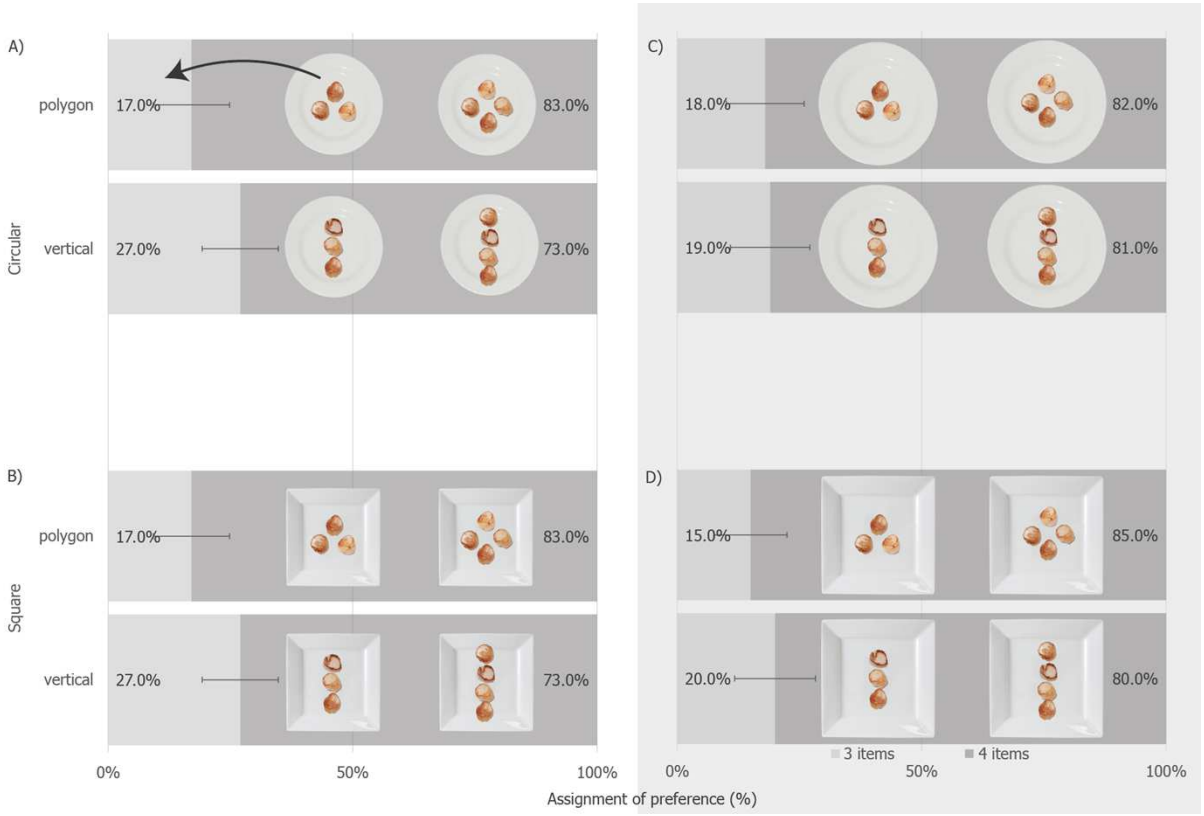


Figure 4. The percentage of people preferring one dish over the over for each of the Plate Arrangement, Plate Size, Food Shape and Food Item cells in Experiment 2 (error bars are 95% CI, all differences p<.001).

DISCUSSION

There was no statistically significant evidence to support the scenario that plate overcrowding influenced dish selection here. At first glance, the results on Experiment 2 are rather different from those of the preceding study. Here, by far the majority of our participants preferred the 4-item dishes, as opposed to the 3-item dish. In Experiment 1, though, the magnitude of this preference was much smaller; indeed, when the items were arranged vertically, participants preferred the 3-item dish over the 4-item dish. It should be noted, though, how the pattern of results in Figure 4A and 3B, which tested participants on the same plate sizes as Experiment 1, if one ignores the magnitude of the preference difference, resembles that seen in Figure 3A and 2B for Experiment 1: when the items were arranged vertically, more participants picked the 3 item dish, relative to

when the items were arranged as a polygon. Given how Experiment 1 found evidence of food arrangement and this study does not, we will continue exploring food arrangement in subsequent studies – it may just be that a ceiling effect here led to this difference between studies.

Why do we observe such a discrepancy between this study and the previous, in terms of *magnitude of preference difference*? One possibility is that the population from which the participants were sampled are quite different to each other, with those in Experiment 1 predominantly coming from the UK (and of the sort who visit science museums), whilst those in this study mostly came from North America; indeed, a potential major driver here could be that North Americans generally have larger meal sizes (as explored in the movie 'Super Size Me' (Spurlock, 2004).

EXPERIMENT 3: equating portion sizes

Here we scaled the 4-item dish so that it contained exactly the same amount of food as the 3-item dish, by factoring in the height of the scallops. By doing so, we factor out the influence of portion size in this study (if we ignore the fact that perceived portion size often differs from actual portion size – as was discussed in the Introduction), which should give us a clear indication whether or not participants prefer one dish over the other for perceived portion size, or for the likely aesthetic difference between 3 or 4 elements being placed on the dish. Again, we hypothesise that participants will prefer the 3-item dish over the 4-item dish. Do note that plate size has been shown to influence perceived portion size (for this and other such influences, see Benton, 2015, and Hollands et al., 2015). However, as we never contrast portions over different sized dishes, such effects should not confound the results presented here.

MATERIALS AND METHODS

Participants

One hundred (31 female) were recruited from Amazon's Mechanical Turk to take part in the experiment in return for a payment of .35 US dollars. The participants ranged in age from 18 to 69 years (M = 33.1 years, SD = 10.9). The experiment was conducted on 10/06/2015, from 16:00 GMT onwards, and over a one-hour period. The participants took an average of 89 seconds (SD =

104) to complete the study. All of the participants provided their informed consent prior to taking part in the study.

Stimuli, Design and Procedure

This study was identical to Experiment 2 except that the scallops were scaled so that each plate contained the same amount of food. In the previous studies, the scallops were held within 50 x 50 pixel boxes, and we assumed that the height that the scallops were off the plate was approximately 2/3 of this measure (33.3 pixels). Thus, on a three-item plate, the scallops were each tightly held within a 250000 voxel box (3 * 50 * 50 * 33.33). The scallops in the four-item plate were scaled along the x, y, and z axes to 90.86% of their original size so that the boxes they were enclosed within also summed up to this value (4 x 45.43 x 45.43 x 30.29).

RESULTS AND DISCUSSION

A log-linear analysis, as defined in Experiment 2, was run using data from this study (the final model's likelihood ratio was $\chi^2(14)=5.23$, $p=.98$). As in the previous study, the model only retained the effect of Items $\chi^2(1)=41.77$, $p<.001$. 4-item dishes (selected 491 times, or 61.38% of the time) were 1.59 times more likely to be preferred more than dishes with 3 items (309 times; Fisher's exact t-test 95% CI were 57.90% and 64.76%).

The results indicate that, in actual fact, the 4-item dishes were preferred over the 3-item dishes, which is contrary to the wide spread claim that that odd-number of items should be preferable. Unfortunately there is a further confound that may have swayed this result: Could portion sizes have been distorted by some other means? We then tested whether our arrangements were thought different in portion size due to potential distortions brought about by psychological illusions of volume perception.

EXPERIMENT 4: scaling study

MATERIALS AND METHODS

Participants

One hundred participants (51 female) were recruited from Amazon's Mechanical Turk to take part in this study in return for a payment of 1 US dollar. The participants ranged in age from 19 to 56 years (M = 30.2 years, SD = 8.02). The experiment was conducted on 5/06/2014, from 14:00 GMT onwards, over a period of three-hours. The participants took an average of 378 seconds (SD = 138) to complete the study. All of the participants provided their informed consent prior to taking part.

Stimuli

The individual scallops used in Experiment 2 and onwards were here dynamically sized, positioned and combined as a dish stimulus as required on each trial (on a plate in most trials; n.b. the plates used were those defined in Experiment 2). The exact scallop images used in each dish that were to be scaled (henceforth termed the 'scaling-dish') were selected randomly, as were their set positions on the plate (care was taken so that the scallops were placed and spaced apart to resemble the vertical and polygonal arrangements that had been used in Experiment 1). The scallops in each dish were simultaneously scaled using the scroll button on the mouse or the left and right cursor keys (where a 'toward the body' scroll and the left cursor key scaled the image downwards) – importantly, the distance between the centre points of the targets did not change on scaling. The minimum size scallops could be scaled so that they tightly fit within a 25 x 25 pixel box. The maximum size was 150 x 150 pixels. The starting size of the scallops was randomly determined but was always such that the scallops fit within a box larger or equal to 40 x 40 pixels and smaller or equal to 60 x 60 pixels.

A target stimulus that was randomly selected from the 5 scallop stimuli was also present on each trial. This stimulus was always sized so that it fit within an 87 x 87 pixel box.

Apparatus

The apparatus varied by participant as the experiment was conducted online. The experiment utilized 'full screen' mode (i.e., utilizing the entirety of the participant's monitor), and took place within a 1024x768-pixel box in the centre of the screen, irrespective of the size of the monitor. The experiment was conducted on the Internet using the Adobe Flash based version of Xperiment (http://www.xperiment.mobi).

Design

A within-participants experimental design was used with all of the participants undertaking all of the experimental trials (trial order was randomised). The dependent variable was the computed scaling factor which the participant applied to the dish of scallops so that they would, together, match the volume of the Target stimulus. A scaling factor of 1 would indicate that the participants scaled the portion so it exactly matched the volume of the target, whilst values smaller than 1 indicate the scallops were sized such that they were of a lesser overall volume than the target. Independent variables were the size of the plate (large or regular), the shape of the plate (circular or rectangular), the number of scallops (three or four) and the arrangement of the scallops (vertical or polygonal). Further trials included dish variants where there were 1 or two scallops only (the latter, arranged vertically or horizontally) and where there was no plate present. Note that there were several further conditions, the data from which will be reported elsewhere (Woods et al, in prep.).

Procedure

On each trial, a screen as shown in Figure 5 was presented. Participants had to scale the size of the scallops shown in the portion of the screen labelled Portion 2, so that they matched the same amount of food as shown in Portion 1. Although Portion 1 was the same size on all trials, the Scallop that was shown as Portion 2 randomly varied in default size over trials. There were 35 trials. At the end of the study we explicitly asked participants “When you did the task, were you resizing Portion 2 so that…”, and offered two choice options “one food item was the same size as Portion 1” , “all the items together in Portion 2 added up to the same amount as in Portion 1.” The 20 participants who chose the first option were excluded from the analysis. There were 35 experimental trials, the data from 20 being reported here.

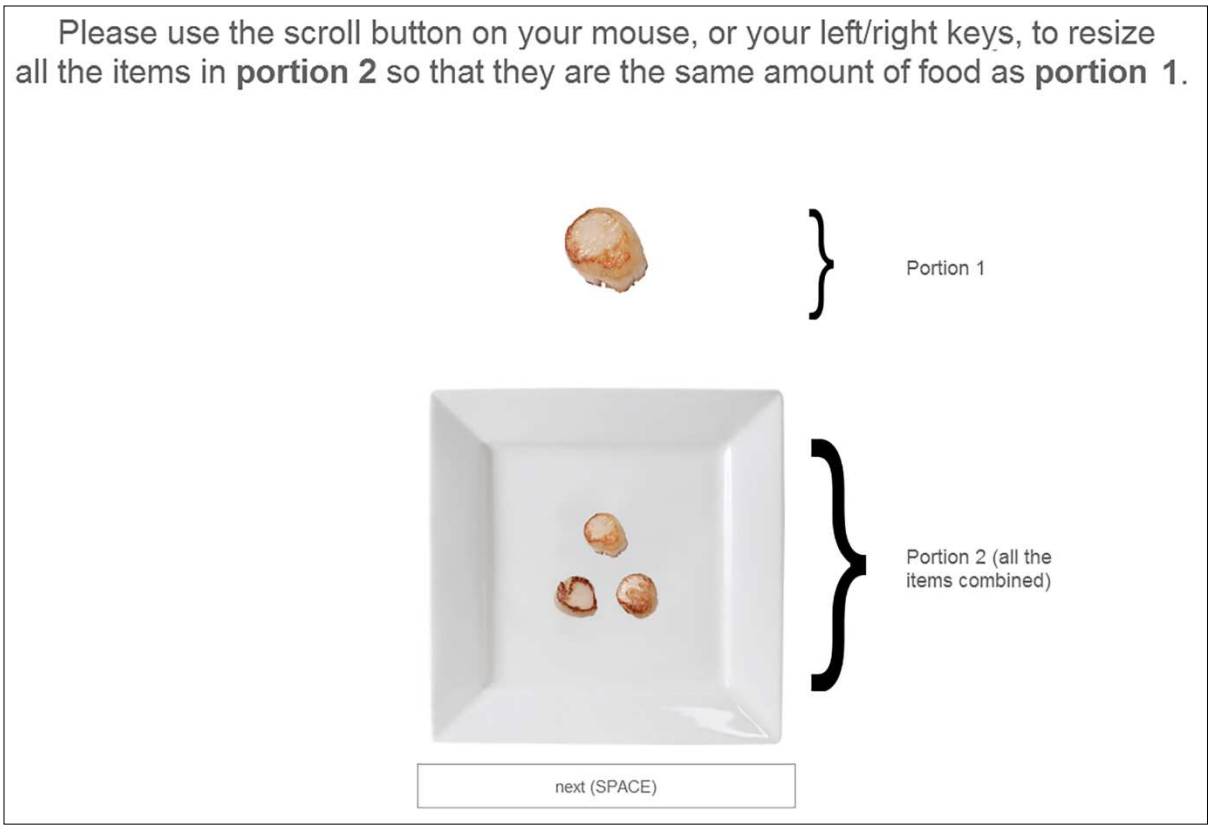


Figure 5: The trial layout, as presented to the participants.

RESULTS AND DISCUSSION

Eleven out of twenty sets comprising the data were not normally distributed $D(80)$, $p<.05$. Log transforming the data mostly corrected this issue, with only one set remaining non-normal, $D(80)=.94$, $p<.001$ (large round plates containing 3 polygonally arranged items). The same set was also was significantly skewed, $p<.001$, and affected by kurtosis, $p<.01$. Another set was also affected by kurtosis, $p<.01$ (regular-sized round plates with 4 vertical items). 0.5% of the scaling data from each dish was found to be outlying (defined as being larger or smaller than the mean +-3 standard deviations) and so was corrected (replaced with the nearest non-outlying data value, mean +-3 standard deviations).

With the majority of the cells of data now being normally distributed, one-sample t-tests were used to test whether the log-scores different from the null hypothesis of that no scaling was required, or log(1), with a Bonferroni corrected alpha threshold set to .05/35 (a further 15 tests on data not

reported here were conducted in Woods et al, in prep). Only data for large round plates with 4 vertical scallops differed significantly $t(79)=3.64$, $p<.001$, requiring scaling of 1.10 to be seen as the same size as the target food. As all other 4-scallop dishes did not so differ (as would be expected given the shift in 3 vs. 4 item preference seen in previous studies), we must assume the null-hypothesis that portion size distortions cannot really account for past findings (that 4-item portions were often preferred over 3-item portions).

Note, though, that in previous research the participants had to choose between 2 dishes, each of which could be differently influenced by scaling factors. Thus, potentially subtler distortions of size (not detectable when contrasting from baseline as done so in the above tests that were essentially *between-participant*), between each pair of dishes, may have driven the shift towards the 4-item dish as opposed to 3-item dish from past studies. To explore this, a 4-way repeated measures ANOVA was conducted on the log scaling data with plate Shape, plate Size, Items and food Arrangement as factors. Items and Arrangement interacted $F(1,79)=22.86$, $p<.001$, $\eta_p^2=.22$, with a posthoc stepwise Newman-Keuls analysis showing that 4-scallops arranged as a polygon requiring more scaling (mean 1.04) than the other conditions (.97; significant main effects that were involved in these interactions are not reported). What this means, in fact, is that the 4-scallop polygon arrangements required were seen as the *smaller* portion than on other dishes (it was required to be scaled by a factor of 1.04, whilst the other dishes by .97, to both be seen as *the same size* as the target portion). We would have expected it to be seen as bigger than the other dishes, for it to explain the apparent 4-item preference over 3-items as seen previously.

To recall the pattern of results from Experiment 1 and Experiment 2, where the preference ratio of 3-item polygonal scallops to 4-item polygonal scallops was greater or more severe than that for vertically arranged items. The fact that here, 4-items are perceived as a smaller portion than 3-items may be linked to this pattern, although at this moment in time, it is unclear how so.

Several other distortions, albeit smaller in magnitude, were also found. Shape and Size also interacted $F(1,79)=5.85$, $p<.018$, $\eta_p^2=.07$, with the same Posthoc procedure revealing that large-round plates required its contents to be scaled more to match the target (mean 1.02) as compared to regular-round (.97), large-square (.99) and regular-square plates (.98). Large-square plates required more scaling than round-regularly sized plates.

A separate repeated measures ANOVA was used to test whether the trials in which there were no plates (only scallops were shown) differed in terms of scaling required to match the target stimulus. The factors of Items and Arrangement were used as factors. There was a main effect of Items, $F(1,79)=8.47$, $p=.01$, $\eta_p^2=.10$, with the 4-item displays (mean 1.04) requiring significantly more scaling than the 3-item displays (1.00).

DISCUSSION

Although some of the portions of food from previous experiments were indeed affected by some distortions of size, there was no systematic effect of distortion of 4-item dishes to appear *larger* than the 3-item dishes, the result of which could be leading participants to prefer 4-items over 3-items.

The tentative conclusion that could now be drawn is that the even number of items on a plate are preferred over an odd number of items. To do so though, one must one ignore several important issues highlighted in the introduction, such as whether 3 vs. 4 items generalise to odd vs. even number of items, as well as whether the effects observed here only are applicable to our scallop stimuli.

COMBINED ANALYSES

The preceding experiments have highlighted the importance of perceived portion size on dish choice, with larger portions tending to be selected over smaller portions. This relationship has been quantified in Figure 6 as a simple-regression model, which shows an extraordinary linear relationship between these factors, for all studies, except for the very first study.

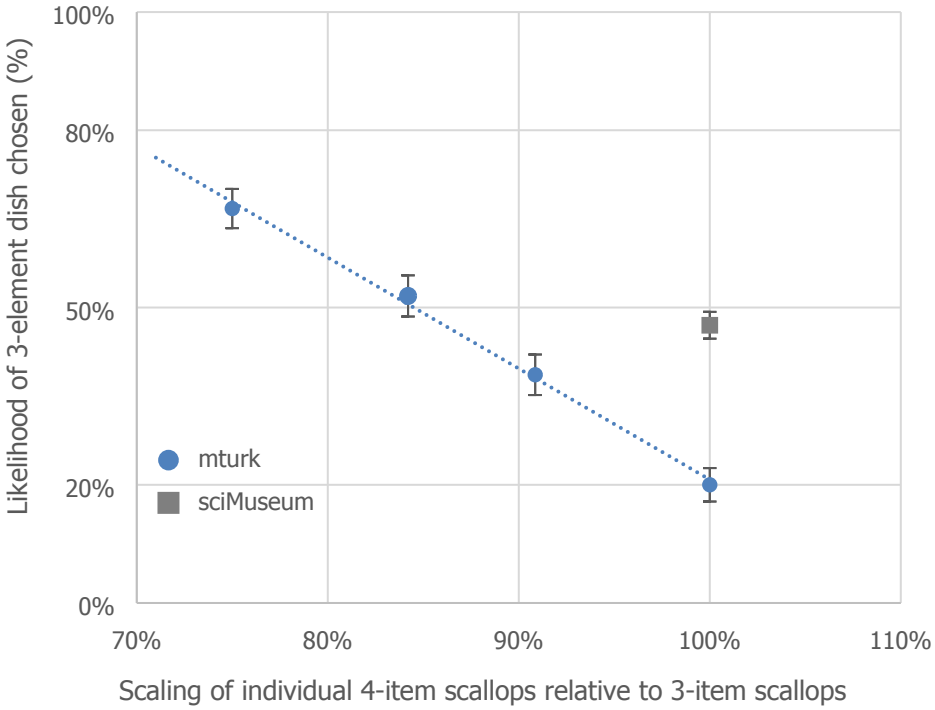


Figure 6: Depiction of the relationship between 4-item scallop scaling and likelihood of 3-scallop dish chosen, over experiments reported so far. Error bars represent the 95% CI derived from separate Fisher's exact binomial tests.

So the question is why the results of the Science Museum study differ so much from the data collected from Mechanical Turk for Experiments 2-5? Recall that the scallops in the original study were not yet scaled to be equal in size in terms of pixels, as done from Experiment 2 onwards. Could the 'fixed' stimuli used in Experiment 1 have led to the above discrepancy? To test for this, we isolated each dish in the study using photo-editing software to estimate total scallop pixels (see Table 1). We then calculated the individual scaling factor present for each condition (square plate x vertical items, .98; square x polygon 1.03; circular x vertical, .93; circular x polygon, 1.04) and plotted this on Figure 7, alongside the ratio of 3-items being selected for each condition.

Table 1: Detailing the size, in pixels, of each scallop that was used in Experiment 1.

Plate shape	Food shape	Food items	pixels	Pixels per scallop

Circle	polygon	3	4606	1711
Circle	polygon	4	6612	1702
Circle	vertical	3	5107	1650
Circle	vertical	4	5919	1653
Square	polygon	3	4545	1535
Square	polygon	4	6427	1607
Square	vertical	3	5133	1480
Square	vertical	4	6598	1515

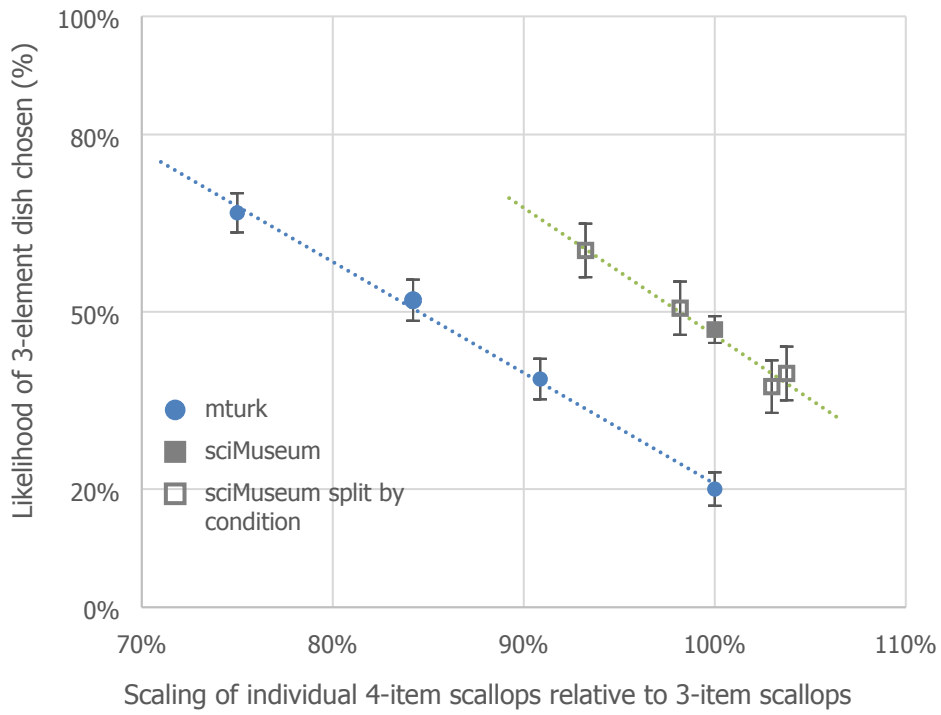


Figure 7: Identical to Figure 6, except that the conditions from Experiment 1 have been added individually as transparent black bordered squares.

Although with 4 data points per model, any inference from statistical analysis must be treated with considerable scepticism, the updated MTurk model's gradient (-.53, 95% CI -.63, -.43) and Science Museum model's gradient (-.45, -.67, -.24) are similar; it is their y-axis intercepts that potentially differ (Experiment 2-5, 111%, 106%, 116%; Science Museum, 121%, 110%, 131%; n.b. overlapping CI).

Why would there be this upward shift of preferring 3-items as opposed to 4-items in the Science Museum study? After another investigation we discovered that the images that were used in the Science Museum study had themselves been arbitrarily scaled by the designers of the citizen science platform so that they were 67.1% smaller in width and height than their original file size (images contained within a 380 pixel x 255 pixel image-file). Furthermore, it also transpired that there were *two* sizes of the original stimuli, with the original images we used being both smaller in scale and held within a differently dimensioned image-file (372 pixels x 306 pixels). By using graphical editing software, we were able to estimate that the Science Museum images were 85.2% smaller in width and height to the images used in Experiments 2-5. Could the difference in overall food size lead to this apparent upward shift between models as seen in Figure 7? We test this hypothesis next. We also tested whether participants' hunger influenced their dish choice.

EXPERIMENT 5: difference due to overall size of stimuli?

In this study, participants undertook a version of the task reported previously where we systematically varied the physical sizes of the dish stimuli on the screen. Both 3- and 4-portion stimuli were resized to the same degree. Note that the monitors of our online participants and thus the stimuli presented differ in terms of size across individuals. To get around this issue, we used a repeated measures design such that all of the participants undertook the trials where differently sized stimuli were presented.

It was hypothesised that if the size of the stimuli was indeed responsible for the difference between the Science Museum study and all of the other studies (as shown in Figure 7), in this study, we should observe a shift in dish preference as we scale the stimuli from smaller to larger in size from that observed for the Science Museum study to that observed for the Mechanical Turk experiments.

We also tested whether the participant's self-reported hunger level influenced the choice design in this task by asking participants how hungry they were.

MATERIALS AND METHODS

One hundred participants (40 female) were recruited from Amazon's Mechanical Turk to take part in the experiment in return for a payment of .50 US dollars. The participants ranged in age from 20 to 67 years (M = 34.8 years, SD = 11.2). The experiment was conducted on 15/06/2015, from 14:30 GMT onwards, and over a 45-minute period. The participants took an average of 105 seconds (SD = 58) to complete the study. All of the participants provided their informed consent prior to taking part in the study.

Stimuli, Apparatus

The stimuli were the same as reported in Experiment 1, except that the scaling of both the 3-item and 4-item dishes (as well as plates) were varied, relative to the original size of the 3-item stimuli as used in Experiment 2. We decided on sizing the stimuli at 100% of those used in Experiment 2 (50 pixels along one dimension), same size of the Science Museum study (42.6 pixels; a difference of 7.39 pixels), smaller than the Science Museum by 7.39 pixels, and larger than the one used in Experiment 2 by 7.39 pixels. In order of size, the stimuli were scaled to 70.44%, 85.22%, 100% and 134.28% of the stimuli used in Experiment 2 and onwards (henceforth termed Small, SciMuseum, Regular, Large).

Design

We used a fully factorial design here with all participants doing all experimental trials. The design was identical to that in Experiment 1, except that an additional factor of plate Size (regular versus large) was included. We also had the participants report their hunger level.

Procedure

The procedure was identical to that used in the studies except that we also assessed participants' self-reported hunger by means of scaled anchored on the left hand side with "not hungry at all" and on the right "very hungry". Hunger scores from this scale varied from 0 to 100.

RESULTS and DISCUSSION

A log-linear analysis was performed, as defined in Experiment 2 but with the additional independent variable of plate Size, using data from this study (the final model's likelihood ratio was $\chi^2(30)=5.12$, $p=1$. The only factor to be retained by the model was Items, $\chi^2(1)=138.91$, $p<.001$, with 4-item dishes (selected 1034 times) 1.83 times more likely to be chosen than 3-item dishes (selected 566 times). The Exact Binomial test 95% confidence intervals for this effect (33.03%, 37.78%) intersected the value predicted by the model for a scaling of 90.86% for the 4-item scallops (37.62%). The lack of any effect of Size indicates that the Small (3-items chosen 33.50% of the time, 95% CI 28.89%, 38.36%), SciMuseum (35.00%, 30.33%, 39.90%), Regular (36.75%, 32.01%, 41.68%) and Large sizes (36.25%, 31.53%, 41.17%) did not significantly differ from each other in terms of the ratio of participants who chose 3-item versus 4-item dishes.

To test whether the hunger level of the participant influenced their dish choice, a correlation was undertaken between the total number of times each participant chose the 4-item dish, and their self-reported hunger score. As the 4-item dish was 1.83 times more likely to be chosen than the 3-item dish (as reported above), we would then expect that, if hunger was an important factor, participants who were more hungry would be more likely to choose the 4-item larger in portion size dish, $r=-.12$, $n=100$, $p=.25$.

There was no evidence that the difference in size between stimuli used in Experiments 2-5 and which were used in Experiment 1 was responsible for their difference in y-axis intercept. There are several possible reasons for this. One possibility is that the within-participants design of this study could have prevented any effects being detectable. For example, consider that the participants here saw many trials one after the other, involving the same task, "which dish do you prefer"? Potentially, after undergoing several such trials, the participants may have 'made up their mind' as to how to respond to each trial (e.g., "I like big portions, so I will always pick the larger portion"), which could sufficiently dilute any normally detectable effects so that they became undetectable. In the Science Museum task, however, the maximum number of trials undertaken by the participants were 2, with the majority of trials thus requiring cognitive effort rather than relying on a quick heuristic.

Another possibility is that the populations from which participants from Experiments 2-6 were sampled from differed in some key criteria from those who undertook the Science Museum experiment. We test this next.

EXPERIMENT 6

A logical step is to rerun the study, but with a different group of participants. Psychology students are well known for being WEIRD (Western, Educated, Industrialised, Rich, and Democratic individuals; see Henrich et al., 2010) and different from Mechanical Turkers (discussed in Woods et al., 2015). Here, we recruited participants from the up-and-coming cloud-sourcing platform Prolific Academic, which actively recruits student participants with no geographic criteria for potential participants, as opposed to MTurk, whose participants are typically North American.

If the difference between the data from the Science Museum reported in Experiment 1, and the rest of the studies reported so far is indeed attributable to some difference over populations, Prolific Academic participants may differ from both these groups too.

To test if this is so, we collect data from stimuli that are sized according to those reported in Experiment 2, 3, and Supplementary Experiments A, and B. We should observe the same gradient as found previously, but with a shift in the y-axis intercept.

MATERIALS AND METHODS

Participants

391 participants (162 female) were recruited from Prolific Academic to take part in the experiment in return for a payment of .35 US dollars. The participants ranged in age from 18 to 67 years (M = 28.4 years, SD = 9.1). The experiment was conducted on 3/07/2015, from 16:00 GMT onwards, over a period of six-hours. The participants took an average of 106 seconds (SD = 48) to complete the study.

Stimuli, Design, Procedure

Identical to Experiment 2, except that Large Plate condition was removed and an additional between participant factor of Scaling was included (how large the 4-item stimuli were, relative to the 3-item stimuli, the levels being 100%, 91%, 84%, and 75%).

RESULTS and DISCUSSION

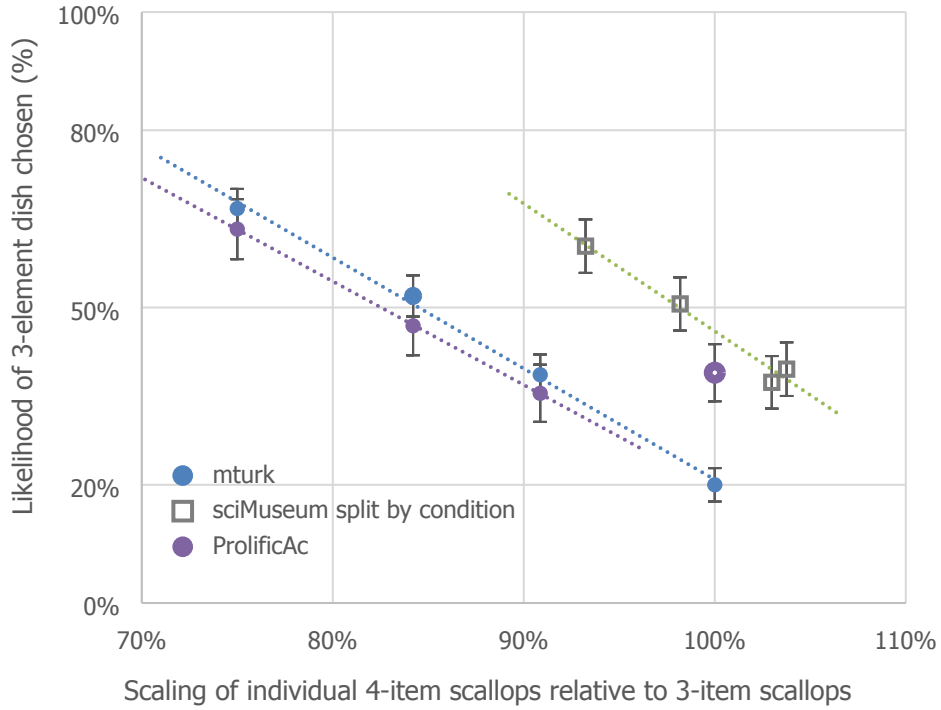
A log-linear analysis was run using data from this study (the final model's likelihood ratio was $\chi^2(14)=3.80$, $p=1$. Items x Scaled x Arrangement interacted $\chi^2(3)=10.54$, $p<.02$. Eight separate Bonferroni corrected Fisher's Exact tests were used to explore this interaction, the results of which are detailed in Table 1.

Table 1: The percentage preference for 3-items relative to 4-items (** $p<.01$, *** $p<.001$, as derived from Bonferroni-corrected exact Fisher's tests; 95% CI in brackets).

	Scaling of 4-item scallops to 3-item scallops			
	100%	91%	84%	75%
Polygon arrangement	30.39*** (24.16, 37.20)	25.27*** (19.20,32.15)	44.33 (37.22,51.62)	62.24** (55.06,69.06)
Vertical arrangement	47.55 (40.53,54.64)	45.70 (38.39,53.15)	49.48 (42.25,56.74)	64.29*** (57.15,70.99)
n	204	186	194	196

The Items x PlateShape interaction was also significant, $\chi^2(1)=4.34$, $p<.05$ with follow-up Exact Fisher tests for each Plate Shape revealing that Square Plates with 4 items (selected 440 times) were selected 1.29 times more frequently than Square Plates with 3-items p<.001 (selected 340 times). There was no such difference for Round Plates (3-items selected 381 times, and 4-items 399 times).

The data for this experiment has been plotted alongside the previous experiments in Figure 8. Whilst the scaled data points for 91%, 84% and 75% form a straight line that does not appear to differ from that of the past MTurk experiments (gradient, -.57, 95% CI -.64, -.5; intercept 111%, 95% CI, 107.53%, 114.56%), the data from the 100% scaled condition unexpectedly does not fit this profile (the transparent purple point in the Figure).



607

608 *Figure 8*: Identical to Figure 7 but with the results of Experiment 6 added. Note that the large transparent
609 purple point did not follow the pattern of the other data points from this study.

610

611 Back in Experiment 2 (labelled E2 in Figure 8) we observed an Arrangement x Items interaction
612 and speculated that this arose due to overcrowding on the plate for the 4-item in relation to the 3-
613 item vertical dishes. One possibility in the current study is that the vertically aligned scallops were
614 likewise seen as overcrowding the plate. For some reason however, the participants here *preferred*
615 this compared to when the items were not so overcrowded, hence the 4-item preference from 3-
616 item preference shifted upward, as shown in the graph.

617 Unfortunately, due to the confound of stimuli sizing for the Science Museum study discussed in
618 the *COMBINED ANALYSES* section, we do not have data for Vertical dishes at this level of 4-item
619 scaling. We do, however, speculate that such an effect would be present, and would increasingly
620 influence the results as overcrowding increased yet further (as indicated by the shaded box in the
621 Figure). As overcrowding is not, however, the focus of this research, we will leave the speculations
622 of the drivers of this finding to future research.

In terms of our initial hypothesis, despite the above unexpected finding, there is little evidence to support the idea differences in terms of population led to the shift in intercept between MTurk studies reported here, and the results of the Science Museum. In the General Discussion, we flesh out reasons why this may be the case.

GENERAL DISCUSSION

There was no support for our original hypothesis that an odd number (3) of items on a plate would be preferred to an even number (4). After controlling for portion size (Experiments 2-3, Supplementary Experiments A and B), testing for plate overcrowding (Experiment 2) and perceptual distortions (Experiment 4), only one group of participants were found to sometimes prefer 3-item dishes as opposed to 4 (Figure 3; Experiment 1, the Science Museum); on contrast, two further groups of participants recruited through MTurk (Experiments 2-5) and Prolific Academic (Experiment 6) preferred 4-item dishes over three. We will discuss several major issues with these findings after briefly summarising each of the experiments in turn.

OVERVIEW OF STUDIES

The results of the first experiment, conducted in collaboration with the Science Museum with 1816 participants, were ambiguous, with 3-items being preferred over 4-items when those items were vertically orientated and on a circular plate only. In all other conditions, 4-items were preferred. This was followed up with a series of experiments that, in turn, tested, and helped control for several confounds, the first of which was ensuring that the individual food items were the same size over conditions (not so in the first study).

Next tested was whether plate overcrowding had influenced findings in the first study. Experiment 2 explored this potential confound by testing whether the ratio between plate size and the surface area covered by the food influenced the plating preference. There was no statistical evidence for such an effect, although, descriptively, effects of food liking were less strong on larger plates than on regular plates, which warrants future research. Unexpectedly, 4 item dishes preferred in all experimental conditions.

Several further experiments tested whether the difference in portion size over conditions in Experiment 1 acted to confound the results. The relative size of the 4-item portion was reduced relative to the 3-item portion in Experiment 3, and in Supplementary Experiments A and B, with the general finding being that the larger the portion, the more people were likely to pick that portion over a smaller portion.

Expectedly, and opposite to our hypothesis, 4-items were preferred over 3 when portion sizes were equated. In Experiment 4, we tested whether there was a perceptual distortion of portion sizes such that the 4-item dish seemed greater in size than the 3-item portion, but there was no real evidence for this. This issue is explored in a complementary paper (Woods et al., in prep).

By means of a combined analysis, there was clear evidence that portion size plays a key role in deciding which plating people prefer, with larger perceived portions more likely to be chosen. Furthermore, we found that quantifying the portion sizes over experimental condition, the Science Museum study, seemed to obey this principle as well. However, whilst the rate of change of the findings over the first study versus other studies seemed equivalent (that is, portion size change tallied with liking change), the scaling at which a 4-item dish required to seem the same size as the 3-item dish differed.

One explanation for this variation was that all the stimuli used in Experiment 1 were actually smaller than those used in the subsequent studies. When explicitly testing for this with a repeated-measures design in Experiment 5 (to get around the issue of hardware variation in online research), this issue was, however not found to influence plating preferences.

Another explanation was that population differences from which Experiment 1 participants were from (the general public in the UK mostly) and those recruited from in other studies (Mechanical Turk) led to this shift. Experiment 6 attempted to test this by recruiting from a third population (Prolific Academic) to see if this population's preferences differed from the other two populations; these individuals though also adhered to the same portion size dish preference principle. This new sample did not really differ from the samples recruited through Mechanical Turk, but nevertheless we cannot rule out that population differences have indeed caused the discussed difference in results. Furthermore, it seemed that plate overcrowding has a different impact on plate preference for this group, than for Mechanical Turkers.

679

680 CAVEATS

681 **Generalizability**

682 Can this research, exploring whether three items of food on a plate (specifically scallops) are
683 preferred to four items, be generalised to 'odd versus even'? A related issue is just how
684 generalizable are the results obtained with the scallop stimuli used here. One way of answering
685 these questions is to survey just how frequently the different number of items appears in natural
686 dining situations (see Michel et al., 2015a, b, for a methodology that could help elucidate this
687 mystery).

688 Another consideration is just how generaliseable these findings are to the general public. Indeed,
689 if we base the choice of 3 vs. 4 items on culturally-based aesthetic preferences (i.e. in China, 4
690 items are regarded as beautiful), it seems even harder to rule out a conclusion. Is this a western
691 chef only issue? Note that the idea that odd items being preferred comes from chefs working in
692 'fine dining' restaurants, where a full meal is served over smaller portions, and hence more
693 attention is paid to the aesthetic appeal of each portion, ruling out, to some extent, the perceived
694 portion-size factors that seem to have been important in determining the findings reported in this
695 paper.

696 **Experimental design**

697 Here, the pairs of dishes presented to participants were always identical in terms of plate shape,
698 plate size and food arrangement (vertical versus polygonal), which meant that it was impossible to
699 test for interactions between these factors. The decision to go with this design was to minimise the
700 number of trials participants would have to undertake (Experiment 2 onwards) or to ensure
701 sufficient numbers of participants per group in the Science Museum study (we were delighted that
702 1816 participants took part in our study and expected a smaller sample size).

703

Conclusions

The evidence reported in this study suggests that an even (4) number of items generally preferred over an odd (3) number of items, even though this might not always be the case. It seems likely that such an effect varies over populations (cf. Experiment 6), though future research will be needed to verify this claim. To come with a clear result on this topic will be challenging, as it will have to control for several major issues pertaining to the experimental design and generalizability of the findings. Several tangential findings arose as a result of teasing apart the initially promising findings in the original study performed at the Science Museum. Although not significant for 3 or 4 items on a plate (Experiment 4), we report evidence in a complementary article that the perceived portion size of vertical and horizontal pairs of items is distorted (Woods et al., in prep). There was also suggestive evidence for the negative impact of plate overcrowding on liking, but, again, this warrants further studies to be verified and properly tested for.

While the plating of food is important in modelling our opinion of, and subsequent enjoyment of a dish of food, the discipline of aesthetic food plating seems to remain more in the realms of artistic sensitivity, to that of an empirical science.

720

721 REFERENCES

722

723 Arnheim R. (1988). *The power of the center: A study of composition in the visual arts*. Berkeley,
724 CA: University of California Press.

725 Bajaj, D. (2013). *Effect of number of food pieces on food selection and consumption in animals*
726 *and humans*. Unpublished doctoral dissertation. Arizona: Arizona State University.

727 Benton, D. (2015). Portion size: What we know and what we need to know. *Critical Reviews in*
728 *Food Science and Nutrition*, **55**, 988-1004.

729 Bouleau C. (1980). *The painter's secret geometry*. New York, NY: Hacker Books.

730 Daneyko, O., Zavagno, D., & Zanuttini, L. (2011). Lightness effects in Delboeuf and Ebbinghaus
731 size-contrast illusions. *Perception*, **40**, 464-473.

732 Deroy, O., Michel, C., Piqueras-Fiszman, B., & Spence, C. (2014). The plating manifesto (I): From
733 decoration to creation. *Flavour*, **3**:6.

734 Deroy, O., & Spence, C. (2014). Can you find the golden ratio in your plate? *Flavour*, **3**:5.

735 Forde, C. G., Almiron-Roig, E., & Brunstrom, J. M. (2015). Expected satiety: Application to
736 weight management and understanding energy selection in humans. *Current Obesity Reports*, **4**,
737 1-10.

738 Geier, A. B., Rozin, P., & Doros, G. (2006). Unit bias a new heuristic that helps explain the effect
739 of portion size on food intake. *Psychological Science*, **17**, 521-525.

740 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral*
741 *and Brain Sciences*, **33**, 61-135.

742 Hollands, G. J., Shemilt, I., Marteau, T. M., Jebb, S. A., Lewis, H. B., Wei, Y., Higgins, J. P. T.,
743 & Ogilvie, D. (2015). Portion, package or tableware size for changing selection and consumption
744 of food, alcohol and tobacco (Review). *The Cochrane Library*, **9**.

745 Kerameas, K., Vartanian, L. R., Herman, C. P., & Polivy, J. (2015). The effect of portion size and
746 unit size on food intake: Unit bias or segmentation effect? *Health Psychology*, **34**, 670-676.

747 Latto, R., Brian, D., & Kelly, B. (2000). An oblique effect in aesthetics: Homage to Mondrian
748 (1872–1944). *Perception*, **29**, 981-987.

749 Latto, R., & Russell-Duff, K. (2002). An oblique effect in the selection of line orientation by
750 twentieth century painters. *Empirical Studies in the Arts*, **20**, 49-60.

751 McClain, A. D., van den Bos, W., Matheson, D., Desai, M., McClure, S. M., & Robinson, T. N.
752 (2014). Visual illusions and plate design: The effects of plate rim widths and rim coloring on
753 perceived food portion size. *International Journal of Obesity*, **38**, 657-662.

754 Michel, C., Velasco, C., Fraemohs, P., & Spence, C. (2015a). Studying the impact of plating and
755 cutlery on ratings of the food served in naturalistic dining contexts. *Appetite,* **90**, 45-50.

756 Michel, C., Velasco, C., Gatti, E., & Spence, C. (2014). A taste of Kandinsky: Assessing the
757 influence of the visual presentation of food on the diner's expectations and experiences. *Flavour*,
758 **3**, 7.

759 Michel, C., Velasco, C., & Spence, C. (2015b). Cutlery matters: heavy cutlery enhances diners'
760 enjoyment of the food served in a realistic dining environment. *Flavour*, **4:**27.Michel, C., Velasco,
761 C., Woods, A. T., & Spence, C. (submitted). On the importance of balance to the aesthetic plating
762 of food. *Appetite.*

763 Michel, C., Woods, A. T., Neuhäeuser, M., Landgraf, A., & Spence, C. (2015c). Orienting the
764 plate: Online study assesses the importance of the orientation in the plating of food. *Food Quality*
765 *& Preference,* **44**, 194-202.

766 Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human
767 preference. *Annual Review of Psychology*, **64**, 77-107.

768 Shimoyama, S. (1976). Translation of *Sakuteiki: The Book of the Garden* (Tokyo: Town and City
769 Planners). Attributed to Toshitsuna Tachibana, late 11th/early 12th Century.

770 Spence, C., & Piqueras-Fiszman, B. (2014). *The perfect meal: The multisensory science of food*
771 *and dining*. Oxford, UK: Wiley-Blackwell.

772 Spence, C., Piqueras-Fiszman, B., Michel, C., & Deroy, O. (2014). Plating manifesto (II): The art
773 and science of plating. *Flavour*, **3**, 4.

774 Spurlock, M. (Producer/Director) (2004). *Super Size Me* [Motion Picture]. United States: The Con.

775 Styler, C., & Lazarus, D. (2006). *Working the plate: The art of food presentation*. New York, NY:
776 John Wiley.

777 Surkys, T., Bertulis, A., & Bulatov, A. (2006). Delboeuf illusion study. *Medicina (Kaunas)*, **42**,
778 673-681.

779 Van Tonder, G. J., & Lyons, M. J. (2005). Visual perception in Japanese rock garden design.
780 *Axiomathes*, **15**, 353-371.

781 Wilkinson, L. L., Hinton, E. C., Fay, S. H., Ferriday, D., Rogers, P. J., & Brunstrom, J. M. (2012).
782 Computer-based assessments of expected satiety predict behavioural measures of portion-size
783 selection and food intake. *Appetite*, **59**, 933-938.

784 Wilson, A., & Chatterjee, A. (2005). The assessment of preference for balance: Introducing a new
785 test. *Empirical Study of the Arts*, **23**, 165-180.

786 Yoshino, D., Kimura, A., & Noguchi, K. (2009). Visual illusion and aesthetic preference: Some
787 common stimulus properties. *Gestalt Theory*, **31**, 29-42.

788 Zampollo, F., Kniffin, K. M., Wansink, B., & Shimizu, M. (2012). Food plating preferences of
789 children: The importance of presentation on desire for diversity. *Acta Paediatrica,* **101***,* 61-66.

790 Zampollo, F., Wansink, B., Kniffin, K. M., Shimuzu, M., & Omori, A. (2012). Looks good enough
791 to eat: How food plating preferences differ across cultures and continents. *Cross Cultural*
792 *Research*, **46**, 31-49.

793 Zellner, D. A. (2015). Effect of visual cues on sensory and hedonic evaluation of food. In A.
794 Hirsch (Ed.), *Nutrition and chemosensation* (pp. 159-174). Boca Raton, FL: CRC Press.

795 Zellner, D. A., Lankford, M., Ambrose, L., & Locher, P. (2010). Art on the plate: Effect of balance
796 and color on attractiveness of, willingness to try and liking for food. *Food Quality and Preference,*
797 **21**, 575-578.

798 Zellner, D. A., Loss, C. R., Zearfoss J., & Remolina S. (2014). It tastes as good as it looks! The
799 effect of food presentation on liking for the flavor of food. *Appetite,* **77C**, 31-35.

800 Zellner, D. A., Siemers, E., Teran, V., Conroy, R., Lankford, M., Agrafiotis, A., Ambrose, L., &
801 Locher, P. (2011). Neatness counts. How plating affects liking for the taste of food. *Appetite,* **57**,
802 642-648.

Supplementary Material

SUPPLEMENTARY EXPERIMENT A: 4 items much smaller in terms of volume than 3 items

The original goal of this study was to scale the 4-item stimuli so that together they would be the same size as the 3-item dish. Due to an error in scaling though, the 4-item food elements were, in fact, much smaller than the 3-item food elements. However, this gives us the opportunity to see whether 'the table turns' as it were, and we now find that 3-item dishes that are larger in portion size than 4-item dishes suddenly become the preferred dish.

MATERIALS AND METHODS

Participants

One hundred participants (38 female) were recruited from Amazon's Mechanical Turk to take part in the experiment in return for a payment of .35 US dollars. The participants ranged in age from 19 to 64 years (M = 32.7 years, SD = 10.3). The experiment was conducted on 9/06/2015, from 16:00 GMT onwards, over a one-hour period. The participants took an average of 95 seconds (SD = 144) to complete the study. All of the participants provided their informed consent prior to taking part in the study. The experiment was reviewed and approved by the Central University Research Ethics Committee at the University.

Stimuli, Design and Procedure

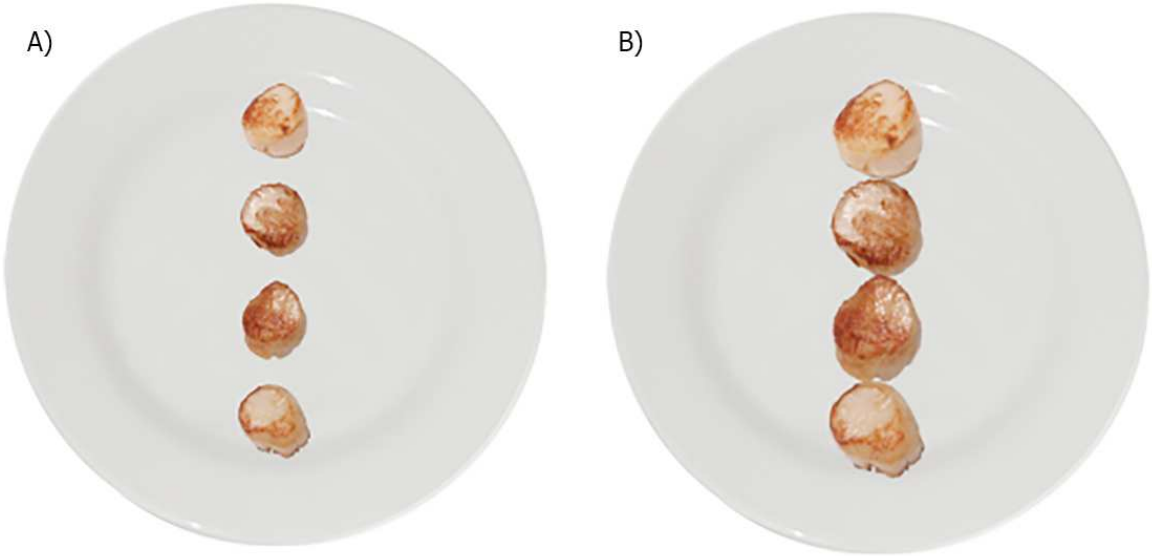
Here, the 4-item stimuli were individually 75% the size (along the vertical and horizontal dimensions) of each of the scallops in the 3-item plate.

RESULTS AND DISCUSSION

A log-linear analysis, as defined in Experiment 2, was conducted on the data from this study (the final model's likelihood ratio was $\chi^2(14)=.97$, p=1). The model only retained the effect of Items,

829 $\chi^2(1)=91.54$, $p<.001$, with 3-item dishes (selected 534 times, or 66.75% of the time) 2.00 times
830 more likely to be preferred as the 4-items dishes (266; Fisher's exact t-test 95% CI were 63.37%
831 and 70.01%).

832 Given that the 3-item dishes were indeed preferred over the 4-item dishes, this implies that portion
833 size plays a role in how people decide between two plates of food, in terms of their preference.
834 Note the absence of any effect of plate size in this study, compared to the Experiment 2. By making
835 the portions smaller, the plates here perhaps appeared less cluttered (or did not seem over-full; see
836 Figure 1). Although we cannot rule out that it was portion size, not visual clutter that led to such
837 effects being observed in Experiment 2 (and speculated on in Experiment 1), drawing inspiration
838 from Occam's razor, the former being the simpler (and explainable) alternative perhaps makes it
839 the most likely.



840

841 Figure 1. The vertical 4-item circular dish from Supplementary Experiment A (A) and Experiment
842 2 (B).

843

SUPPLEMENTARY EXPERIMENT B: what happens at when the plates contain the perceptually equivalent amounts of food?

We expected that when the 3- and 4-item dishes were identical in terms of the perceived amount of food, that other effects besides 'number of scallops' would be easier to detect. In order to estimate the scaling of the 4-item scallops to use, we created a linear model to predict the likelihood of participants deciding upon the 3-item dish (as opposed to 4-item dish) based on the scaling of 4-item scallops relative to 3-item scallops from Experiments 2-5 (we did not include Experiment 1 given the remaining ambiguity concerning why it differed from the other experiments so far; plotted in Figure 1). Using this model, the scaling to use for the 4-item scallops so that participants would be equally like to decide on a 3- or 4-item dish was estimated to be .84.

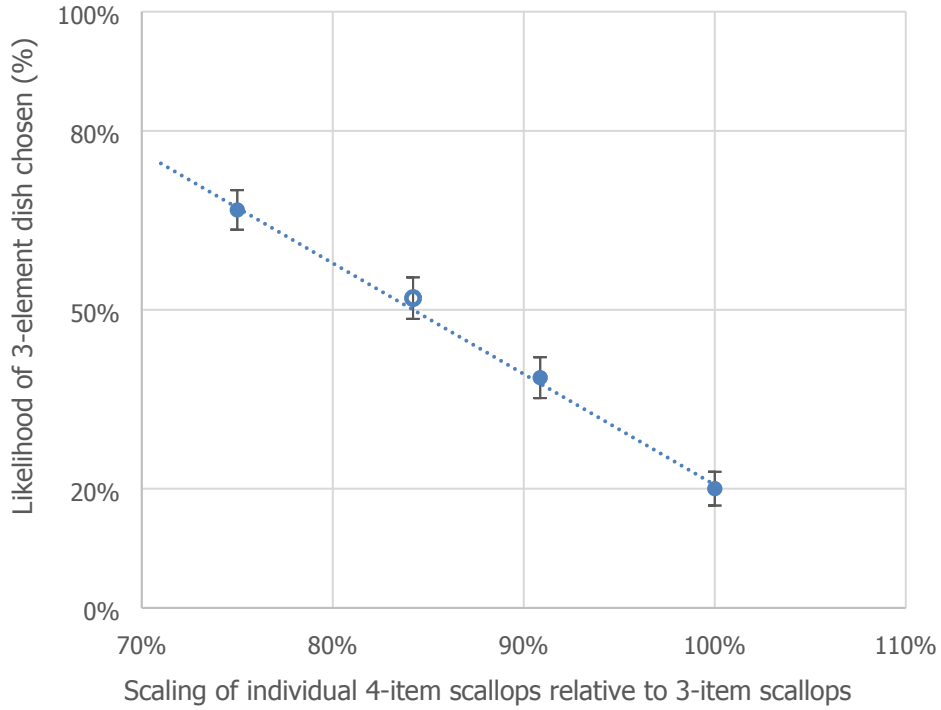


Figure 1: Figure depicting the relationship between 4-item scallop scaling and likelihood of the 3-scallop dish being chosen, over Experiments 2 and 3, and also from the Supplementary Experiment A. The results of this study, Supplementary Experimental B is also shown (the transparent blue circle). Error bars represent the 95% CI derived from separated Fisher's exact binomial tests.

MATERIALS AND METHODS

Participants

One hundred and two participants (34 female) were recruited from Amazon's Mechanical Turk to take part in the experiment in return for a payment of .35 US dollars. The participants ranged in age between 19 and 56 years (M = 32.7 years, SD = 9.3). The experiment was conducted on 12/06/2015, from 16:00 GMT onwards, over a one-hour period. The participants took an average of 95 seconds (SD = 109) to complete the study. All of the participants provided their informed consent prior to taking part in the study. The experiment was reviewed and approved by the Central University Research Ethics Committee at the University.

Stimuli Design and Procedure

Were the same as in Experiment 2 with the exception that scaling of the food elements was such that the 3-item and 4-item food portions were perceptually equal.

RESULTS AND DISCUSSION

A log-linear analysis, as defined in Experiment 2, was conducted using data from this study (the final model's likelihood ratio was $\chi^2(15)=3.62$, $p=1$). There were no significant effects remaining in the model and only its constant was retained. There was thus no evidence that factors other than 'number of food items' played any role in dish selection in this particular study.