Repeatability of odour preferences across time†

S. Craig Roberts,a,* Jan Havlíčekb and Marion Petrie[c]

ABSTRACT: Human body odour is often associated with negative attributions, hence the term ‘malodour’. Another perspective is that odours contain biologically meaningful information involved in communication of social cues, notably in perception of suitable mates. This evolutionarily informed perspective indicates that we retain capacity to infer mate quality through olfaction (e.g. preferring odours of high-quality or genetically compatible individuals). From either perspective, knowing the extent to which body odour is stable over time is important: either in order to fully understand how perfumes might interact with body odour or whether the biological cues gained from odour are reliable. In addition, from the second perspective, odour-based mate preferences should also be relatively stable over time, especially if both traits and preferences are genetically influenced. Here we measured repeatability in young women of body odour preferences for male odours, over a 3-month period. We also compare stability of body odour preferences with that of preferences for faces and fragrances. We find that preferences for all stimuli were highly repeatable over time. Since the odour stimuli used were repeated samples from the same set of men, repeatability of preferences also indicates odour constancy of individuals over time. Our results on both odour constancy and repeatability of preference have implications for the perfume industry and also lend weight to the assumption that body odour constitutes a meaningful cue of quality that can be used in individual assessment during human interactions. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: olfaction; smell; attractiveness; body odour; mate choice

Introduction

There is growing research interest in the perception of human body odour. One perspective sees body odour as an undesirable trait laden with negative associations,[1] for example, it may be associated with poor hygiene, leading to attribution of low social status[2] and promotion of outgroup prejudice such as racist sentiment.[3] Negative associations such as these lead to the term ‘malodour’ and use of personal hygiene products including deodorants to remove it, as well as use of perfumes to mask and replace it. As Stoddart[4] put it: ‘human beings behave as if they are afraid of smelling like human beings, for human beings smell bad’.

However, another approach sees body odour in a very different light: as a meaningful way of perceiving biologically relevant information about other individuals during social interaction. This approach views human body odour as a means of social communication in much the same way that a variety of odours are used in animals,[5–7] notably how odour functions in signalling aspects of mate quality to potential mates.[8,9] Studies in this area measure preferences of individuals for odours of opposite-sex individuals and interpret emergent preferences in terms of the fitness benefits that might result (for reviews, see Roberts and Little,[10] and Havlíček and Roberts[11]). Recent examples from the literature include the findings that odor may enable choice of genetically compatible partners,[11–13] or reflect other aspects of a male’s quality,[14,15] or of a woman’s position within the menstrual cycle.[16–18] From this perspective, attempts to remove, mask or replace body odour could interfere with the detection of this social information, with potential disruption of mate choice and other social attributions,[19] although recent studies indicate that preferences and choices of perfumes may actually be sensitive to this, such that people choose perfumes that complement or perhaps even enhance their own body odour.[20–22]

From either of these perspectives, the extent to which an individual’s body odour is stable or labile is an important piece of information. For example, it could influence perfume design and patterns of consumer purchasing behaviour. However, it might be especially important to know whether body odour is very changeable from the second perspective. This is because, if perfumes are indeed selected to suit an individual’s odour, then the degree and rate of change might influence alteration in preference for the odour. From a theoretical viewpoint, it is also important because adaptive explanations for mate preferences or other social cueing effects predict that odours should be relatively stable over time, particularly if there is a genetic component to odour quality.

In addition, this second perspective assumes that variation in preference for traits should be relatively stable.[23,24] Although preferences for opposite-sex physical traits (which includes body odour) may vary between individuals dependent on aspects of their perceived social status,[10,25] and may vary within individuals depending on physiological state (e.g. menstrual
cycle or use of hormone contraceptives\textsuperscript{[12,26–27]}, the extent to which human preferences for body odours are stable over time has not been tested. In this paper, we test the repeatability of women's preferences among the body odours of six male odour donors over a period of about 3 months (i.e. spanning three menstrual cycles). We hold other factors constant, including only women who were not using oral contraceptives, and testing them during the fertile phase of the menstrual cycle. We also compare the stability of body odour preferences against the stability of preferences for a different olfactory stimulus, using an equivalent number of fragrance ingredients. Finally, we also compare it with another physical trait used in mate assessment in another sensory modality (judgements of facial attractiveness from photographs).

**Methods**

**Participants**

Experimental participants were staff or students at the University of Newcastle, UK. Sixty-three heterosexual women, aged 18–32 years, contributed to the study as raters (not all women completed all tasks so sample sizes for presented analyses vary slightly, between 59 and 63). As menstrual cycle stage can influence both odour sensitivity and perception\textsuperscript{[28,29]} and facial preferences,\textsuperscript{[30,31]} we controlled for this by testing women only in the late follicular stage (days 10–14) of their cycle (using forward counting\textsuperscript{[26,27]}), Ninety-two men (aged 18–31 years) also took part, forming a pool of potential participants for inclusion in stimulus sets (i.e. provided odour samples and facial photos) for different women.

Women participants judged the attractiveness of male odours, faces, and perfume ingredients, on two occasions. The second tests were carried out during the third follicular phase after the first test, producing an inter-test interval of approximately 3 months.

**Stimuli Collection**

Odours were collected by asking men to wear a cotton T-shirt, which we provided, for two nights in bed. Participants were instructed to avoid using perfumed products (e.g. deodorants, aftershaves, perfumed soaps) for 2 days before, and while they wore the shirts. They were provided with non-perfumed soap (Simple; Accanta Health & Beauty Ltd, Birmingham, UK) for washing. Men were also asked to avoid certain strong-smelling foods (e.g. garlic, blue cheese), to avoid excessive alcohol consumption and smoky bars, and to sleep alone in their beds. All men were non-smokers. These measures were designed as far as possible to capture underlying male body odour, while avoiding other confounding variables, as done in previous studies\textsuperscript{[12,27,32–34]}. On the second morning, participants returned the shirts within labelled plastic bags. Shirts were halved (to increase sample numbers, being cut from throat to navel) and then immediately frozen at $–30^\circ\text{C}$. Freezing does not affect perceptual quality of body odours.\textsuperscript{[26,35]} Men contributed shirts on request over a number of months. Shirts used in these experiments were normally worn within 10 days prior to the testing session; in all cases, women smelled shirts worn by men approximately 3 months apart.

Digital colour photographs of the men's faces were also taken under standard lighting conditions. Men were instructed to look directly at the camera and adopt a neutral expression. Following common procedure in testing for facial attractiveness,\textsuperscript{[10,36]} images were digitally masked so that only the face was visible, to avoid confounding effects of hairstyle and clothing.

**Body Odour Preference Testing**

We presented odours of six males to each female. Both male and female participants were taking part in a larger study\textsuperscript{[26]} on possible associations between odour preferences and genotype at the major histocompatibility complex (MHC), following an experimental protocol based on earlier MHC studies by Wedekind and colleagues.\textsuperscript{[12,13]} For this reason, the set of six male odour donors chosen for any particular female was non-random, but balanced, with respect to MHC-genotype (three MHC-similar and three MHC-dissimilar men). This form of selection for inclusion in stimulus sets should have no influence on the ability to test repeatability of preferences. However, it does have the advantage of producing unique combinations of odour donors for each woman, thus avoiding the potential situation where results could be confounded by choice of stimulus set (e.g. a set of six where a small number were extremely pleasant or unpleasant to all smellers and thus artificially creating congruence in judgements; cf. Wedekind\textsuperscript{[37]}).

Two hours before female participants arrived for testing, the six selected shirts were removed from the freezer and placed in clean glass jars at room temperature. Jars were sealed with aluminium foil. When participants arrived, jars were inverted and shaken vigorously, and a triangular nose-hole was cut in the foil to allow odour assessment. Women were informed that they were about to smell the odours of six men, and were asked to take a brief sniff of each jar in turn, to gain some idea of the range of odours available. They then smelt each jar more carefully, and, using pen and paper, rated each odour on a seven-point scale (where 1 = not at all pleasant, 7 = extremely pleasant). They were asked to rate scores against the specific questions: "How pleasant do you find this smell?" (the same question as used by Wedekind and colleagues\textsuperscript{[12,13]}), and "Is this how you would like your long-term partner to smell?" (cf. Roberts et al\textsuperscript{[26]}, hereafter we term this as ‘desirability’). Finally, women were asked to re-arrange jars in order of preference, and these rank orders were recorded.

Order of odour presentation in the first test was determined by the level of MHC similarity between women smellers and male odour donors, alternating between MHC-similar and MHC-dissimilar (i.e. jars 1, 3 and 5 being MHC-similar or dissimilar). In the second test, the order of jars was reversed so that any similar scoring or ranking could not be accounted for by order of presentation. Participants were pre-informed that the odours were from the same men that they had smelled in the first test, but in a different order.

Although odour donors were the same in both tests, the T-shirts used were worn by the donors on different nights, approximately 3 months apart (and were discarded after use). Differences in body odour quality which arise due to variable conditions of odour donors would tend to act against repeatability in our experiment. Thus our design is conservative with respect to repeatability but allows us to assess more realistically the repeatability of preferences for the same men's odours across time. We also assessed repeatability using a constant odour stimulus, by using fragrance ingredients (see the section ‘Perfume preferences’), and this...
Flavour Fragr. J. 2013, 28, 245–250

Repeatability of odour preferences across time

enabled us to distinguish effects of changing body odour and changing odour preference.

Perfume Preferences

We assessed repeatability of odour preference for a constant odour stimulus by asking women to also rate their preference among perfume ingredients. This task was scheduled at the end of the test session, after assessment of body odours. Women smelled 15 perfume ingredients, again as part of a larger study. For the purpose of this study, to provide an equivalent comparison with preferences for the six body odours and faces, we used responses to the first six of these ingredients. These fragrances were essential oils of bergamot, lime, lemon grass, myrrh, camphor and clove, presented as three droplets of undiluted oils in unmarked, brown vials. Women rated how much they would like these odours as potential ingredients to be included as a perfume for themselves (again, on a scale of 1–7, where 1 = not at all, 7 = very much).

Facial Preferences

After women had completed the odour assessments, they were asked to view and score male images on a liquid crystal display computer screen. Similar to the odour tasks, women rated attractiveness of these faces in answer to two questions (‘How attractive do you think this man for a short-/long-term relationship?’, where short-term was contextualised by examples, such as attractiveness for a dinner date or holiday romance) using a seven-point rating scale, according to how attractive they thought the faces were (where 1 = very unattractive, 7 = very attractive).

Women saw the images of the six men who had been used in the odour preference task, but in different order. Again, each woman saw unique sets of male faces. In the second test, the same images were used, but as with the odour presentations, image order was also altered (with a different order than the odour presentations). Subjects were also asked to inform the experimenters if any faces were familiar. This occurred on four occasions (one image for each of four women) and these were omitted from the analyses.

Analyses

Repeatabilities were calculated by carrying out Spearman’s rank correlations on scores, for each woman, given to odours or faces in the first and second tests. The observed median correlation coefficients were then tested against the hypothesised medians (in each case, zero, which would represent no effect) using one-sample Wilcoxon signed-rank tests. Differences in repeatability were assessed cross-modally using related-samples Wilcoxon tests on the correlation coefficients between the two test sessions for each task.

Results

Odour Preferences

Mean pleasantness scores across women for the six male body odour samples ranged from 2.33 to 5.5 (mean = 3.78, standard error = 0.07). Women scored the six male odours consistently across the two tests (Figure 1). The median Spearman’s correlation coefficient was 0.35 (range, –0.72 to 0.98) for judgements of odour pleasantness. The frequency distribution of within-individual correlation coefficients was negatively skewed and significantly different from chance (i.e. zero; z = 3.97, P < 0.001; Figure 1a). Similar results were found for ratings of desirability (median correlation coefficient = 0.18; range, –0.58 to 0.98, z = 3.67, P < 0.001; Figure 1b).

Correlations between preferences were similar when odours were ranked in order of potential partner preference by the women, as a separate task from the ratings (Figure 2). Using ranks of odours, the median correlation between the two tests was 0.34 (range, –0.77 to 1) and this was significantly different from chance (z = 3.92, P < 0.001).

There was also high repeatability of ratings of perfume ingredients across the two tests. The median Spearman’s correlation coefficient was 0.77 (range, –0.31 to 1) and significantly different from chance (z = 6.83, P < 0.0001; Figure 3). There was significant difference amongst the fragrances (F5,378 = 20.14, P < 0.0001); taking the first test scores, bergamot was rated as most pleasant (mean ± SE, 3.75 ± 0.21), followed by lime (2.86 ± 0.22), then camphor (2.44 ± 0.17), lemon grass (1.89 ± 0.17), clove (1.81 ± 0.17) and myrrh (1.67 ± 0.11).

Facial Preferences

Facial attractiveness scores given to male faces in each of the two tests were also positively correlated. When faces were rated for short-term attractiveness (perceptually most similar to odour pleasantness ratings), the median Spearman’s rank correlation coefficient was 0.67 (range, –0.29 to 0.96) for judgements of facial attractiveness. As found in the odour judgements, the frequency distribution of coefficients was negatively skewed and significantly different from chance (z = 6.52, P < 0.001; Figure 1c). The medians for rating long-term attractiveness also differed from chance (M = 0.60, z = 6.60, P < 0.001; Figure 1d).

Cross-modal Comparisons

Among the male-derived stimuli, individual repeatability in ratings across time was specific to sensory modality. To test this, we calculated the correlation coefficient, for each woman, between her ratings from the first and second rating session of each type of stimulus (e.g. body odour), giving us between 59 and 63 correlation coefficients per stimulus type. We then took (for example) the set of correlation coefficients from the body odour desirability ratings, and set up a new correlation between these two sets, reporting the resulting correlation coefficients in Table 1. There were significant positive correlations between the different body odour ratings, such that women who scored similarly (i.e. gave relatively high or low scores to the same individuals) across the two time periods for one odour judgement also did so for the other judgment. Similarly, the short-term and long-term facial judgments were significantly correlated. However, there was no relationship between repeatability in odour ratings and repeatability in facial ratings. Repeatability was higher for ratings of images than ratings of body odour (short-term face attractiveness vs. odour pleasantness, z = 3.43, P = 0.001; long-term face attractiveness vs. odour desirability, z = 3.79, P = 0.001). Repeatability in perfume ratings was not significantly correlated with that for either the facial or body odour judgments.
Repeatability for fragrance ratings were higher than that for either body odour pleasantness ($z = 4.11$, $P < 0.001$) or desirability ($z = 4.81$, $P < 0.001$). However, although there was no difference between repeatability in rating faces or perfume ingredients (short-term face attractiveness vs. fragrance, $z = 1.41$, $P = 0.158$; long-term face attractiveness vs. fragrance, $z = 1.86$, $P = 0.063$), across-test correlations were slightly higher for ratings of fragrances than facial images.

**Discussion**

The main aim of this paper was to test repeatability of preferences for body odours across time. We found that women’s relative preferences for six male odours (and faces) were similar when measured in preference tests spaced 3 months apart. Our results indicate that qualitative assessments of relative attractiveness through preference testing in the laboratory are repeatable over at least this interval and thus lend weight to the interpretation of odour and facial judgements as biologically meaningful cues used in mate choice, and potentially other attributions used in social perception.

We found significant repeatability in measured preferences for body odours, even though the odour stimuli were not identical in the two tests. These body odour samples originated from the same men and were collected under the same controlled conditions for use in either test, providing a more plausible test of ‘real-world’ repeatability for body odour preferences than if samples were stored in a freezer between tests and reused. Repeatable judgments that we report here are thus likely to be partly determined by constancies in body odour over time[38] which would indicate underlying genetic contributions to individual body odours. These are known, for example, from twin studies of similarity in odour chemistry[39] and perception of these odours,[40] and from studies investigating similarity in odours among kin.[32,41,42] These constancies are robust enough for measurable consistency in women’s relative preferences among the six male odour donors, despite potential change in the quality of male body odour across the 3-month period. Such change could occur as a result of change across sampling times in any environmental variable that influences body odour. This could include change in health[43,44] or grooming habits[45] but perhaps most likely would come from dietary differences immediately before or during sampling (for a review see Havlíček.
perception of axillary odour. Although our participants were asked to avoid eating highly aromatic foods such as those rich in garlic around the time of sampling, experiments finding changes in odour after ingesting certain foods show a likely route for how variation in odour across tests might arise.

Comparisons between the repeatability of body odour preference and preference for the perfume ingredients enable us to assess likely repeatability in preference for a constant odour stimulus in a mate choice context. Repeatability in preference for the six perfume ingredients was very high, and significantly higher than consistency for body odours, indicating that women’s preferences may be less labile than suggested by results from the body odour ratings. In other words, although women’s relative preference among the same men over time is repeatable even allowing for environmental influence on the men’s odour, it would be even higher had we used more controlled body odour stimuli (i.e. the same samples which had been frozen between tests). Alternatively, higher repeatability for the perfume ingredients could be due to high distinctiveness between ingredients and thus higher correlations, because some odours (e.g. bergamot) were consistently rated as more pleasant than others (e.g. myrrh).

Consistency in body odour across time, and in its perception by potential partners, is a critical prerequisite for ideas about odour playing a role in social interactions, particularly in influencing women’s mate preference.[50] and also carries implications for the design of perfumes. Recent work indicates that perfume preferences may be at least partly genetically determined[51,52] and that individual choice of perfume complements underlying body odour to create a more attractive emergent odour blend.[20–22] Our results showing consistency of body odour, body odour preferences and fragrance preference all suggest that perfume choices are likely to be similarly consistent over time.

Comparisons of the levels of repeatability of preference across the different stimuli revealed two further surprising results. First, individual variation in repeatability was task-specific: women who produced highly repeatable ratings for body odour (for example) did not produce highly repeatable ratings of faces or fragrances. We might have expected greater consistency in the two olfactory tasks (body odour, fragrance) or in the two mate choice tasks (body odour, faces), but this appeared not to be the case. At least for the two olfactory tasks, these results would appear to rule out individual variation in olfactory perception (e.g. sensitivity, discrimination) as an explanation for the observed pattern. Women therefore appear to vary in their attention towards different kinds of stimuli, and this variation is relatively consistent across time. Second, although not statistically significant, repeatability tended to be slightly higher for fragrance perception than for judgements of facial attractiveness. This could suggest that, even for judgments of faces, women’s preferences vary slightly over time, perhaps according to variation in psychological attributes such as self-perceived attractiveness which influence facial judgements.[53] This variation was unlikely in this case to have included change associated with menstrual cycle phase, as all women were tested each time in the fertile phase of their cycles; however, because we did not confirm ovulation using ovulation tests or similar methods, this remains a possible cause of some of the variation. At the very least, it indicates that olfactory judgements can be just as repeatable as those made in the visual domain, at least in a mating context. This again has implications for the perfume industry, as consistency in liking for one fragrance or another

**Figure 2.** Frequency distribution of Spearman’s rank correlation coefficients for ranking by pleasantness of six male odours by women in two tests spaced three months apart. Rankings were more consistent than would be expected by chance ($P < 0.001$)

**Figure 3.** Frequency distribution of Spearman’s rank correlation coefficients for rating of pleasantness of six fragrance ingredients by women in two tests spaced three months apart. Rankings were more consistent than would be expected by chance ($P < 0.001$)

and Saxton.[46]) For example, perceptual differences in body odour arise through experimental manipulation of dietary regimes in animals (e.g. protein in rodents[47]). In humans, experimental manipulation of meat intake[48] or garlic[49] similarly influences
is informative about consumer decision-making processes. Choosing a perfume is a complex, multi-dimensional and idiosyncratic decision, involving consideration of price and marketing materials as well as the perceptual qualities of the perfume itself. With regard to the latter, our results demonstrate that repeatability across time seems to be high. Taken together, our results therefore suggest that once consumers find a perfume that suits them, fidelity to it is likely to be high.

In summary, a significant degree of repeatability persists for odour preferences, indicating both that individual odour quality is relatively robust with respect to environmental influence and that preferences based on body odour are relatively stable. Furthermore, in circumstances where body odours are less environmentally influenced, repeatability could potentially approach or equal that found in other sensory modalities. Such temporal consistency of body odours is an important prerequisite for effective functioning in signalling of status and individuality, and also carries fundamental implications for the fragrance industry.

Acknowledgements

We thank all our volunteers for taking part, and Andreas Natsch and two anonymous reviewers for their comments. The work was funded by the Wellcome Trust (grant number 058394) to Marion Petrie. Jan Havlícek is supported by the Czech Science Foundation grant (P407/10/1303) and Charles University Research Centre (UNCE 204004).

References