

Reproducible research into human semiochemical cues and pheromones: learning from psychology's renaissance

5

Tristram D. Wyatt ^{1,2}

¹ Centre for Biodiversity and Environment Research, University College London, London, UK

² Department of Zoology, University of Oxford, Oxford, UK

10

Corresponding Author:

Tristram D. Wyatt ¹

Centre for Biodiversity and Environment Research, University College London, London, UK

Email address: t.wyatt@ucl.ac.uk

15

Abstract

As with other mammals, smell in the form of semiochemicals is likely to influence the behaviour of humans, as olfactory cues to emotions, health, and mate choice. A subset of semiochemicals, pheromones, chemical signals within a species, have been identified in many mammal species. As mammals, we may have pheromones too. Sadly, the story of molecules claimed to be ‘putative human pheromones’ is a classic example of bad science carried out by good scientists. Much of human semiochemicals research including work on ‘human pheromones’ and olfactory cues comes within the field of psychology. Thus, the research is highly likely to be affected by the ‘reproducibility crisis’ in psychology and other life sciences. Psychology researchers have responded with proposals to enable better, more reliable science, with an emphasis on enhancing reproducibility. A key change is the adoption of study pre-registration which will also reduce publication bias. Human semiochemicals research would benefit from adopting these proposals.

1. Introduction

Our sense of smell is undoubtedly an important influence on our behaviour, a view supported by ample evidence in the papers in this special issue [xx] and previously reviewed by Doty (1981), Schaal and Porter (1991), Wyatt (2014) and McGann (2017). Our olfactory sense is well developed, even if it has often been underestimated.

In 2019, we celebrated the 60th anniversary of the first chemical identification of a pheromone, the silkmoth female sex pheromone, and with it a new word, pheromone (Wyatt 2009, 2017). This is also an important moment for the study of human semiochemicals as we could benefit greatly by applying the ideas rapidly emerging from psychology to improve the reproducibility and reliability of our conclusions (Bishop 2019; Munafò et al. 2017; Nelson et al. 2018). As I will explore in this paper, there are good reasons to doubt the most studied ‘putative human pheromones’, androstadienone and estratetraenol, despite their popularity with experimenters (Doty 2014; Wyatt 2015). The results are quite possibly explained by false positives. Similarly, much of the human semiochemical literature exploring cues, such as the odours of fear, is characterized by small scale experiments: across psychology, typically only 50-70 % of such experiments can be replicated (Section §5). How might that apply to our own field? In the second half of this paper I will explore constructive ways forward that are emerging from what has been called ‘psychology’s renaissance’ (Section §6)(Chambers 2017; Munafò et al. 2017; Nelson et al. 2018). I should say at the outset that I offer these suggestions respectfully, well aware that I am not an experimental scientist in this field. I admire the scientists who tackle the hard topic of human semiochemicals. As I am not competing with such scientists for grants, I hope I can offer a broad and relatively disinterested perspective.

I start by briefly making the case that mammals use semiochemicals, including pheromones, and that as mammals, we might too.

2. Other mammals rely on semiochemicals (humans surely do too)

Smells are very important to mammals (humans included) but few odour molecules are pheromones, most are cues. Pheromones occur superimposed on the background of hundreds of molecules making up a highly variable odour profile which differs between individuals. These variable odours come from many sources, including those secreted by the organism, from bacteria, picked up from conspecifics, and from diet (Natsch this volume; Wyatt 2014)pp 2-16, 284-291.

(a) Cues to health, individual identity, mate choice, and physiological state

Animals may use some molecules in this complex chemical odour profile as cues to assess health or physiological state of other conspecifics, much as a mosquito uses carbon dioxide emitted by its mammal host as a cue to locate it. The mammal does not release carbon dioxide in its breath as an evolved signal to attract mosquitos, but the insect has evolved to respond to this stimulus.

There is some evidence that animals avoid sick conspecifics (Wyatt 2014)pp77-79. Humans may be able to detect and avoid infected individuals (Olsson this volume). Different infections and other diseases produce their own characteristic smells (Shirasu & Touhara 2011). One possibility is that animals, including humans, have evolved specific responses to these smells of infection. Perhaps more likely is that the evolved response is a more general one: to detect and avoid a conspecific that is not 'smelling right', unlike other conspecifics. This would be more generalizable for any new infection not previously met in the population.

The highly variable chemical profile which differs between individuals can be learnt as an odour 'fingerprint' (Wyatt 2010, 2014). This can be used to recognise siblings, neighbours, or mates(s).. The memory of the individual odours of siblings when growing up together, may be used as a cue to avoid these kin as mates when adult. One of the contributions to the individual differences in odour profiles comes from the immune system (MHC), by mechanisms still not fully understood. MHC-related odours may be learnt and used to avoid kin matings in mice. The evidence for MHC (HLA) influences on mate choice in humans may not be as clear as once thought (Havlicek this volume; Lobmaier *et al.* 2018).

Learning can be programmed, as in the case of imprinting a memory of the lamb's chemical profile by the mother sheep (ewe) by release of neurotransmitters in her brain, triggered by stretching of the vaginal wall/cervix during birth (Wyatt 2014)pp 215-221. To my knowledge, there is no evidence of this occurring in human mothers though it might (Schaal this volume).

Physiological changes in individuals may be reflected in the odour molecules they give off, providing cues for conspecifics to gauge the internal state of others. For humans, experiments suggest that the odours of fear or stress may be detected as cues (de Groot this volume; Pause this volume). Some preliminary studies have explored the effect of sniffing axillary (armpit) sweat from individuals watching sexually arousing films (e.g.(Zhou & Chen 2008)). Whether

these are different cues from those elicited by fear has not been directly tested so far as I know: both situations, sexual or fearful, would lead to emotional release from armpit apocrine glands (Natsch this volume; Wilke *et al.* 2007). There's some evidence that human males may be able to detect changing odours in women's ovulatory cycles (Haselton & Gildersleeve 2016; Havlíček *et al.* 2006).

(b) Mammal pheromones

Pheromones are chemical signals which have evolved with this function in the signaller (Wyatt 2014, 2017). Pheromones occur superimposed on the background of the highly variable chemical profile which differs between individuals. A pheromone such as the male mouse sex pheromone darcin (major urinary protein MUP20) is the same in the urine of every dominant male mouse; it does not distinguish any one male from other male mice (it is the other variable molecules in his profile that do).

Over evolutionary time, some olfactory cues can evolve into pheromone signals, presumably if they are consistent and reliable indicators of physiological state for example (Wyatt 2014)pp18-21. This has been the suggested route for the evolution of fish sex pheromones, many of which are related to hormones which would have been circulating in the blood and then leaking out in urine or across the gills. Over evolutionary time, these molecules have been modified, are produced as pheromone signals at particular periods related to courtship, and have specific receptors in the receiver's olfactory system (Sorensen & Wisenden 2015). Alternatively, other molecules may become pheromones but the production of these is under hormonal control. So, castrated male mice do not produce darcin as this is under androgen control; only dominant male mice produce darcin in their urine (Hurst & Beynon 2013). Subordinates do not, but will if they become the dominant. Other examples are discussed in Wyatt (2014).

The steps needed to establish that a molecule(s) are a pheromone are basically unchanged since the first was identified by Butenandt and colleagues in the 1950s, using the wing fluttering of the male silkmoth as the bioassay for presence of the female sex pheromone (Wyatt 2014, 2017). These steps start with demonstrating a behavioural or physiological response (quantified in a repeatable experiment called a bioassay) which is mediated by a potential chemical stimulus such as a secretion; then isolate, identify, and synthesize the bioactive molecule(s); then confirm that the proposed molecule(s) at natural concentrations are necessary and sufficient to recreate the original response with the original bioassay (Wyatt 2014, p49 ff). These steps or their equivalent, including rigorous bioassays, remain an essential part of pheromone identification today. Without the publication in full of this systematic approach, no claim that a molecule or combination of molecules is a pheromone is credible (Wyatt 2017). I have proposed an operational definition of pheromones to specify the minimum pragmatic evidence needed to demonstrate a pheromone (Wyatt 2014, 2017).

Pheromones have been identified in many mammal species: for example, in mice from darcin in male urine and ESP1 in the tears of males, and various other small molecule pheromones such as

(methylthio)methanethiol (Ishii & Touhara 2018), in rabbits, the mammary pheromone 2-methylbut-2-enal (Schaal *et al.* 2003), in goats, 4-ethyloctanal as part of a male pheromone with a primer effect activating the gonadotropin releasing hormone (GnRH) pulse generator in females (Murata *et al.* 2014). There is still much to learn, with the possibility in the goat, as the authors leave open, that as the single molecule they propose does not completely reproduce the full activity of the natural chemical stimulus from the male, the full pheromone may include other molecules. Even for long-studied phenomena such as the acceleration of puberty in juvenile female mice by male urine ('Vandenbergh effect'), we are still unsure which molecules, large and/or small, are really involved (Flanagan *et al.* 2011).

Either or both of the mammalian olfactory systems may be used to detect pheromones, depending on the pheromone and species. In mice, large molecule pheromones such as ESP1 and darcin are detected by V2R receptors in the vomeronasal organ (VNO) while small molecule pheromones are detected by both the VNO and main olfactory epithelium (MOE) (Ishii & Touhara 2018). Pheromones in rabbits and sheep are detected by the MOE so the lack of a functional VNO in humans does not rule out pheromones (Wyatt 2014).

Identifying pheromones remains challenging in vertebrates and, perhaps even more than in other mammals, we should expect humans to present problems. A first difficulty in mammals generally is the large number of background molecules in their highly variable individual profile. Subtractive techniques to reveal molecules against this background that are present in, for example, in all intact males but not castrated males, is one approach to identify candidate molecules for further bioassay (Murata *et al.* 2014; Schwende *et al.* 1986). A second difficulty is the development of a reliable and repeatable bioassay which is also biologically relevant to the response being investigated. An indication of the challenge is the continuing mystery of the identity of the female dog pheromone, whose existence was noted by the ancient Greeks. Recent Polish work has shown how hard it is to design a bioassay: out of context, even the urine of oestrus females is not attractive and, as is often the case, sniffing durations are not helpful, and female dog models do not work (Jezierski *et al.* 2019) (incidentally this paper shows the value of publishing negative results).

A further challenge is that although pheromones, by definition, elicit stereotyped behavioural and/or physiological responses, in all animals including mammals, these responses are modulated by context, time of day, and many other factors including the receiver's genetics, age, sex, hormonal state, dominance status, and recent experience (Wyatt 2014). In arguing that mammals do not have pheromones, Doty (2014), contrasts the variable responses of mammals with the supposedly unchanging automatic responses of insects to pheromones. However, insect responses are modulated too. For example, male *Agrotis* moths do not respond to female sex pheromone for up to 24 hours after mating (see Wyatt (2014) for this and other examples). Even though sex pheromone still stimulates the olfactory sensory neurons in the male moth's antennae, the brain response is blocked until enough time has elapsed for the male to have replenished his accessory protein stores to go with his sperm. Male hamsters only respond to female pheromone

if they are well fed, giving them sufficiently high blood testosterone levels in their hypothalamus. The response to pheromones may be affected by memory of individual odours (Wyatt 2014)pp218-221. For example, in female sheep and goats, the primer ‘male effect’ from the male’s pheromone seems to be more effective if the male is novel, likely detected by his individual odour (Jorre De St Jorre *et al.* 2014).

Caution is needed when using other species’ pheromones as evidence for the likelihood that pheromones of a particular kind exist in a second species. For example, while a mammary pheromone guides pups to the nipple in rabbits (Schaal & Al Aïn 2014), in mice it seems that mouse pups respond to the signature odour of their mother learned in utero (Logan *et al.* 2012). The use of a male tear-gland protein pheromone, ESP1, in mice is no indication that tears will be important in human semiochemical communication. A particular molecule being a pheromone in one species does not rule it out being a pheromone (or a component of one) in another species but being a pheromone in one species does not make it more likely that the same molecule is a pheromone in another species. For example, androstenone and androstenol are thought to be the male pig sex pheromone (though see (Doty 2014)) but the presence in small quantities in human armpits is no evidence that these are human pheromones (Doty 2014) (next section).

3. Putative human pheromones

There have been three waves of ‘putative human pheromones’, starting with ‘copulins’ in the 1970s, androstenone and androstenol in the 1980-1990s, and from 1991, androstadienone and estratetraenol (reviewed in Doty 2014; Havlíček *et al.* 2010; Wyatt 2015; Wysocki & Preti 2004). (No molecules have been proposed for the different subject of human menstrual synchrony and the phenomenon itself has been questioned (Doty 2014)).

The literature on androstadienone and estratetraenol starts with the paper by Monti-Bloch and Grosser (1991) in a conference proceedings sponsored by the EROX Corporation, which was patenting the two molecules as ‘putative human pheromones’ (Wyatt 2015). As explored in Wyatt (2015), no information was given in Monti-Bloch and Grosser (1991), nor in any patents, about how these molecules were found and shown to be pheromones – we don’t even know which parts of the body the original samples came from.

The Monti-Bloch and Grosser (1991) paper was not one that *could* have established that the proposed molecules were pheromones. It was not designed to. It basically only reported the test of 5 molecules, all supplied without further justification, by the EROX Corporation. The samples were only tested on 20 men and 20 women. Further problematic was that the recordings were from ‘VNO’ tissue when subsequent research has concluded that human adults do not have a functioning VNO (Smith *et al.* 2014). There is no paper anywhere which provides any evidence, meeting the defining steps outlined in Section §2b, that androstadienone and estratetraenol are pheromones.

There is no more reason to use these particular molecules than any other pair of random molecules from the human armpit. There is no more reason now than in the year 2000 when

Jacob and McClintock (2000) reported a study using the molecules, using them only because of Monti-Bloch and Grosser (1991).

In the almost 30 years since 1991, there have been some 50 experimental studies using androstadienone and/or estratetraenol. The studies have shown almost uniformly positive results, though sometimes contradictory. The research rationale for using the molecules still seems to be that since so many papers ‘got a positive result’, there must be something there. As explored in Section §5, positive results, including physiological measures, can be baseless. Only since 2015 have a few negative results started to be published (Ferdenzi *et al.* 2016; Hare *et al.* 2017).

We should be asking ourselves: Would we risk the next 20-30 years trying any two other molecules found in human armpits (or any other secretions), chosen at random without solid evidence of efficacy?

4. Human semiochemical research as a branch of psychology

Much of human semiochemicals research including work on ‘human pheromones’ and olfactory cues comes within the field of psychology, whether it is psychophysics or experimental observations of people responding to odours from different sources such as armpit secretions collected from people under stress. While some studies do appear in journals such as *Psychological Science*, many other studies are published in a wide range of journals, such as *Chemical Senses* and *Hormones and Behavior*, that we might not immediately consider psychological. However, the kinds of experiments and questions being investigated tend to fall within the realm of psychology. For example, here are some typical titles:

‘Putative human pheromone androstadienone attunes the mind specifically to emotional information.’ (Hummer & McClintock 2009)

‘Chemosignals communicate human emotions.’ (de Groot *et al.* 2012)

‘A putative human pheromone, androstadienone, increases cooperation between men.’ (Huoviala & Rantala 2013)

This is important because it means that the contemporary debates about how to make psychology more rigorous and reproducible can readily inform the ways we could transform the study of human semiochemicals (Sections §5-7). Like the rest of psychology, the literature on the ‘putative human pheromones’ was overwhelmingly positive, until recently. This fits with the pattern in psychology where it is rare for a hypothesis to be rejected in the published literature.

5. Psychology’s ‘reproducibility crisis’

Much has been written about psychology’s ‘reproducibility crisis’ which came to a head in 2010-2012 (Chambers 2017; Nelson *et al.* 2018; Nuzzo 2015; Yong 2012). Among the events was Doyen *et al.* (2012)’s failure to replicate the results of a famous and highly cited ‘text-book’

social psychology experiment: Bargh *et al.* (1996)'s finding that unconsciously priming young people with words associated with elderly people made them walk more slowly.

However, the reproducibility problem is much wider than psychology. It has been shown, for example, in drug discovery, with the biotech company Amgen unable to replicate more than 6 out of 53 landmark studies in oncology and haematology (Begley & Ellis 2012), translational biomedical research e.g.(Curran 2018), and animal behaviour e.g (Wang *et al.* 2018). Some non-reproducible pre-clinical papers have generated a whole field of study, with hundreds of secondary publications that expanded on ideas in the original study but which did not test its fundamental basis (Begley & Ellis 2012), in an echo perhaps of 'putative human pheromones'.

Were the conspicuous replication failures of single psychological studies typical of the field as a whole? To explore the reproducibility of psychology more generally, a collaboration of hundreds of scientists worked together to replicate 100 of the most important studies published across three leading psychology journals (Open Science Collaboration 2015). Only 40% of the original studies were judged to have been replicated (though see Nelson *et al.* (2018) for a suggestion that a higher proportion might be interpreted as succeeding). Whereas 97% of the original studies had significant results ($P < 0.05$), only 36% had in the replications. Mean effect sizes were halved.

Perhaps this was a feature more common in 'lower impact', more specialist journals? Not so. Camerer *et al.* (2018) evaluated the replicability of 21 social science experiments that were published in *Nature* and *Science* between 2010 and 2015. Only 66% of the original studies could be replicated (with a significant effect in the same direction as originally published). In this large consortium exercise, in many cases involving the cooperation of the original researchers, and despite much larger sample sizes in the experiments, the effect size was uniformly reduced.

These kinds of results, including another major multisite replication study (Many Labs 2) which could replicate only half of the original 28 studies (Klein *et al.* 2018), led science journalist Ed Yong (2018) to write 'Ironically enough, it seems that one of the most reliable findings in psychology is that only half of psychological studies can be successfully repeated'.

(a) Why is science by good scientists going so wrong?

'... many researchers persist in working in a way almost guaranteed not to deliver meaningful results. They ride with what I refer to as the four horsemen of the reproducibility apocalypse: publication bias, low statistical power, *P*-value hacking and HARKing (hypothesizing after results are known).' Dorothy Bishop (2019)

Good scientists doing the best research they can, in good faith, have nonetheless created a situation of unreliability. This is in large part because of a research culture which, under the pressures of publication, survival and grant-getting, has adopted and rewarded research practices (including the four horsemen above) now increasingly recognized as 'questionable' (Bishop 2019; Chambers 2017; Munafò *et al.* 2017; Nelson *et al.* 2018). I will briefly summarize the problems here but for more detailed accounts see, for example, short papers by Munafò *et al.*

(2017) and Nelson *et al.* (2018). For a longer, entertaining and practical discussion see the book *The Seven Deadly Sins of Psychology* by Chris Chambers (2017).

The first problem is publication bias. Novel, ‘statistically significant’, exciting, and seemingly ‘clean’ results with a ‘good story’ are more likely to be published, especially in the most competitive journals. We internalise this, so we only submit such papers (the unsubmitted studies stay in the file drawer). The result in psychology, and much of the life sciences, is a literature which reports positive results in 95% of published papers – the null hypothesis is almost always rejected (Chambers 2017; Munafò *et al.* 2017).

The second is small experiments with low statistical power, with little chance of finding a nominally statistically significant finding that actually reflects a true effect and an exaggerated estimate when a true effect is discovered (Button *et al.* 2013). Instead there is a high chance of false positives – much much higher than the ‘1 in 20’ that a $p < 0.05$ significance threshold suggests, especially when combined with flexible statistical analysis (*p*-hacking)(below).

The third is high researcher degrees of freedom in the search for significance. As Simmons *et al.* (2011) say, “undisclosed flexibility in data collection and analysis allows presenting anything as significant.” This includes flexibility in when to stop collecting data (with peeking at intervals). ‘*P*-hacking’ or ‘data dredging’ is conducting many analyses on the same dataset and just reporting those that were statistically significant but not disclosing these multiple comparisons.

By moderately *p*-hacking two real experiments Simmons *et al.* (2011) demonstrated how easy it is to obtain statistically significant evidence for a transparently false hypothesis: that listening to a Beatles song can change a person’s age (Nelson *et al.* 2018; Simmons *et al.* 2011). (For a demonstration of how easy, try the web app *P*-hacker (<http://shinyapps.org/apps/p-hacker/>)).

These procedures almost guarantee finding some comparison(s) ‘significant’ but make any conclusions highly likely to be false-positives (Simmons *et al.* 2011, 2018b).

The forth horseman, HARKing (hypothesizing after the results are known), is looking back at the data, seeing an exciting pattern, and falsely arguing that it was the *a priori* hypothesis question being tested from the beginning (Kerr 1998). It will feel fine to the experimenter as we have a powerful hindsight bias (Chambers 2017; Munafò *et al.* 2017). It’s been likened to firing an arrow and then drawing a target circle round the arrow after it has landed – you can’t miss. HARKing pretends exploratory research is confirmatory research to test a hypothesis declared in advance. All ‘hypothesis testing’ research is at risk of HARKing or *p*-hacking if the analysis is not specified before the experiment starts (Section §7).

Nelson *et al.* (2018)p514 now suspect that *p*-hacking explains the paradox of how the overwhelming majority of published findings in psychology are statistically significant, despite the overwhelming majority of studies being underpowered and thus unlikely to obtain results that are statistically significant. They suggest that it is failed analyses, not studies, that go into file-drawers and instead, with *p*-hacking, ‘most failed studies are not *missing*. They are published in our journals, masquerading as successes.’

Techniques including funnel-plots to address p-hacking in past research are discussed by Nelson *et al.* (2018). To evaluate the proportion of true effects and indications of likely *p*-hacking in a given set of studies, a technique called *p*-curve (Simonsohn *et al.* 2014, 2019) plots the distribution of reported *p*-values. A '*p*-hacking bump' just below $p < 0.05$ may indicate attempts to get just under the 'significance' line. However, *p*-curve analysis can give an unjustified 'all ok'. With heterogenous sets of published papers, a lack of '*p*-hacking bump' and a right-skewed *p*-curve 'clean bill of health' is not conclusive evidence that there is no *p*-hacking or that the studies have evidential value (see e.g. (Bishop & Thompson 2016)).

The question about *p*-curve analysis illustrates a problem with meta-analysis in a field like psychology where different teams study different effects and studies are very different, even on a single concept (Nelson *et al.* 2018). The meta-analyst cannot reasonably assess if the original results in each paper were based on errors in data collection, design, or undisclosed flexibility in analysis. The end result of a meta-analysis is only as strong as the weakest studies, and meta-analysis can have its own biases not limited to which studies to include, both factors giving the risk that meta-analysis exacerbates rather than solves the problems (Ioannidis 2010; Lakens *et al.* 2016; Nelson *et al.* 2018) p527-8.

The lack of direct replication of experiments in psychology, and across the life-sciences in general, is a key underlying cultural problem. A direct replication seeks to test the repeatability of a previous finding by duplicating the methodology as exactly as possible (Chambers 2017). Whereas physics researchers expect direct replication before accepting new ideas, in psychology for every 1,000 papers only two are a direct replication and only one of these will be by a different lab (Chambers 2017) p50. Without direct replications, psychology has lost the ability to self-correct. Instead of direct replications, psychology replicates concepts with novel experiments that test a related (but different) idea using a different method (Chambers 2017) pp13-16, 48-55, testing for example social priming of stereotypes in different contexts and situations. Arguably, psychology's 'conceptual replications' are not real replications. The original study, which may be erroneous, is not re-tested. Like the rest of psychology, 'conceptual replications' may be the norm in human semiochemical research.

(b) Case history: The rise and fall of 'power posing'

If the 'putative human pheromone' molecules are highly unlikely to have any effect, based on the lack of *any* primary evidence that these are pheromones (Section §3), how is it that some 50 papers using the molecules report positive results? Could these be 'false positives' and the biased reporting only of positive results?

There are precedents in mainstream psychology for a widely believed phenomenon to be shown to probably be based on false positives. One example is 'power posing', apparently supported by more than 50 papers showing positive results that has ultimately been found to be baseless (Section §7a, (Jonas *et al.* 2017)). Power posing is an interesting and attractive idea which has been publicised in a TED talk viewed more than 50 million times: if our confidence is displayed

by our physical posture, could our posture affect our behaviour and physiology? Would posing our bodies into a ‘confident’ ‘powerful’ pose become ‘embodied’ and change the way we behave? In the original study (Carney *et al.* 2010), 42 participants were randomly assigned to be posed by the experimenter for two minutes in ‘open’ expansive ‘high-power poses’, limbs widespread, or instead ‘closed’, contractive ‘low-power’ poses. Participants given ‘high-power’ poses apparently had raised saliva testosterone levels and lowered stress hormone cortisol levels, and were more willing to take financial risks in a gambling task (Carney *et al.* 2010). Participants also self-reported that they felt more powerful (presented as a check for the manipulation) but the key claim of the paper was that “embodiment extends beyond mere thinking and feeling, to physiology and subsequent behavioral choices” (Carney *et al.* 2010).

The first problems came with a much higher powered conceptual replication study, with 200 participants, which failed to show any significant effects of power posing on hormonal levels or risk taking behaviour (Ranehill *et al.* 2015). Self-reported ‘felt feelings’ of power did replicate.

Carney *et al.* (2015) responded in the same issue of the journal with a tabulation of 33 apparently successful studies, including Carney *et al.* (2010), which they concluded supported the hypothesis.

Simmons and Simonsohn (2017) then did a *p*-curve analysis of the 33 studies, which found an average effect size of zero, suggesting that selective reporting was likely to be the reason for the uniformly positive results in the literature.

Cuddy *et al.* (2018) later responded with a *p*-curve analysis of their own including additional studies (published up to December 2016), suggesting that these supported self-reported ‘feelings of power’ (not the original claims by Carney *et al.* (2010) of changes to physiological measures of hormone levels and risk taking behaviour). However, this *p*-curve analysis itself is likely to be unreliable as its conclusions rely on four outliers with unlikely and extreme low probabilities (Simmons *et al.* 2018a).

Courageously, Carney, first author on the 2010 paper, has now stated that she does not believe ‘power pose’ effects are real, listing problems with the original paper, including *p*-hacking thought acceptable at the time (Carney 2016). As important, Carney has helped other researchers conduct a rigorous pre-registered set of replications (Section §7b) that conclude that power posing has no effects on behaviour and physiology, the original 2010 claims, even if it does give a feeling of ‘felt power’.

My reason for covering this story at length is that it is a good example of how a positive literature can uniformly report a non-existent effect. The power posing story shows it is plausible that a positive literature for the ‘putative human pheromone’ molecules could build up without any underlying real phenomenon. It also demonstrates the complexity of teasing out what has been shown and what has not.

6. Doing it better: Psychology's 'credibility revolution' and 'renaissance'

The response of psychology over the last 7 or so years has been swift, innovative, and persuasive. Vazire (2018) argues that rather than describing an ongoing 'reproducibility crisis', with the implication that there are no solutions, she suggests instead a 'credibility revolution', emphasising the increased reliability and solidity of science that can result from the new approaches outlined below. Nelson *et al.* (2018) similarly argue for 'psychology's renaissance'. The science will be better for it – even though it might seem to take longer, with fewer, bigger studies instead of quicker 'exciting', but ultimately unreliable, small studies.

In 'A Manifesto for Reproducible Science' aimed at researchers in human behaviour, Munafò (2017) succinctly bring together proposals including improving ways to remove bias from data collection and analysis: one key proposal is registered reports, described in the next section. They also argue that institutional, funder and other stakeholders need to change incentives for researchers so better practice is rewarded: success should come from producing rigorous, transparent, reproducible science not the opposite. Open science, with sharing of data and software code, is recommended as one solution to small underpowered studies by lone researchers. The Open Science Foundation (www.osf.io) is among the platforms offering many tools to facilitate these new approaches.

7. Registered reports

Registered reports have a time-stamped peer-reviewed protocol agreed before the results of the research are known, and usually before the data are collected (Stage 1 in Figure 1)(Chambers 2019a; Nosek *et al.* 2018). After incorporation of suggestions from the peer-reviewers (which could include statistical advice when it matters most, before data collection), the journal offers an In-principle acceptance (IPA) (Figure 1). Publication decisions at Stage 1 are based on the importance of the research question and quality of the methods proposed to answer that question. With their IPA, the individual researcher can collect their data confident in the knowledge that their publication is effectively guaranteed, whether or not the results are 'significant', so long as the pre-approved protocol is followed (Stage 2 peer review). Pre-registration thwarts p-hacking, HARKing, and publication bias. Exploratory experiments are still encouraged but are labelled as such and treated as preliminary results which would need a pre-registered experiment to test the idea (Nosek *et al.* 2018). It is also possible to pre-register a study protocol with a site such as the Open Science Framework (OSF) (Hardwicke & Ioannidis 2018; Nosek *et al.* 2018).

Figure 1 is the image <https://tinyurl.com/Chambers-Fig> in (Chambers 2019a) <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2019.01299.x>

Caption: Schematic of the registered reports process

Registered reports started with one journal *Cortex*, but 6 years on, more than 200 science journals, including *Royal Society Open Science* and *Nature Human Behaviour*, have adopted them (Chambers 2019a). Funders and journals are starting to partner to peer review proposals together and then offering funding and IPA acceptance in the journal to successful proposals (Munafò 2017). This has the potential to greatly reduce the burden on reviewers as well as improving the design and conduct of experiments.

(a) How well are registered reports working

To date, almost 200 registered report articles have been published. The results are encouraging (Chambers 2019a). As hoped, it seems positive publication bias is greatly lessened: in one survey, 60% of registered reports, both replications and novel experiments, published negative results (a null finding) in marked contrast to the typical 10% in equivalent traditional papers (Allen & Mehler 2019). Researchers will be reassured that registered reports are cited, on average, at or above the impact factors of the journals they are published (tinyurl.com/RR-citations) (Chambers 2019a).

But it is early days and there are things to improve. Hardwicke and Ioannidis (2018) surveyed registered reports to February 2018. One problem they identified is the hidden nature of most agreed protocols so the final registered report could not be compared with the protocol agreed at IPA; advocates agreed that transparent publishing of agreed protocols is essential (Chambers & Mellor 2018). Claesen *et al.* (2019) identified a different problem, undeclared deviations from the preregistration plans of all articles in *Psychological Science* badged as preregistration studies in their 2015-2018 time period. The pre-registration community is characterized by a desire for transparency so ideas for better practice are likely to be adopted.

(b) Power posing revisited: resolved by pre-registered replications

Returning to the question of power posing (Section §5b), in 2017 a special issue of *Comprehensive Results in Social Psychology* was devoted to seven pre-registered studies including direct replications of power posing (all seven studies received advice from one of the original researchers, Dana Carney) (Cesario *et al.* 2017). The seven studies concluded that the original Carney *et al.* (2010) claims of the effects of poses on endocrine levels and risk-taking behaviour did not replicate (Jonas *et al.* 2017). The authors commented on the helpful nature of the pre-registration peer-review. For me, it is a powerful example of a knotty controversy resolved by pre-registered studies.

8. Reproducibility in human semiochemical research

Other areas of science are learning from psychology's response to the 'reproducibility crisis'. For example, cognitive neuroscience researchers recently discussed the application of registered reports and other improved practices in a special forum in the journal *Cortex*, with an introduction by Chambers (2019b). Might we convene a similar discussion among researchers in human semiochemicals, perhaps in a journal such as *Chemical Senses*?

Perhaps one of the most surprising conclusions from psychology's 'reproducibility crisis' is that our routine good practice precautions of randomization and running experiments double-blind are essential but are not enough to ensure the removal of bias. As explored in Section §5a, gathering the data is just the beginning of potential experimenter choices in analysis that can lead to bias.

I think the most significant changes to address the problems would be the adoption of pre-registration as well other proposals outlined in Munafò *et al.* (2017) to tackle cognitive biases. These changes are in addition to the best practices related specifically to the collection, storage, and delivery of semiochemical stimuli (see other papers in this symposium and (Doty 2015; Drea *et al.* 2013; see e.g. Drea 2015; Frumin & Sobel 2013; Keller & Vosshall 2004; Kjeldmand *et al.* 2011; Lapid & Hummel 2013; Lundström *et al.* 2010; Miyazawa *et al.* 2009). The reliability of some of the techniques used in human semiochemistry research, such as neuroimaging including fMRI (Gorgolewski & Poldrack 2016; Poldrack *et al.* 2017), may also deserve further scrutiny. Given the degree of analytical flexibility with the interpretation of fMRI results, pre-registration of the analysis is especially important (Gorgolewski & Poldrack 2016; Poldrack *et al.* 2017).

(a) Re-assessing our existing literature

'that researchers engage in *p*-hacking and that *p*-hacking makes it trivially easy to accumulate significant evidence for a false hypothesis—opened psychologists' eyes to the fact that many published findings, and even whole literatures, could be false positive' (Nelson *et al.* 2018)p513

We need to be more careful with what we cite in our introductions and discussions. We cannot take for granted that the conclusions of papers that we traditionally cite about different aspects of human semiochemistry are reliable. When I have been assessing papers for inclusion in my book chapters or other literature reviews, though I check for controls and whether the study was done blind, I realize that I tend to rely on an assumption that the peer review system has worked, particularly if it appears in a high impact journal. Sadly, as Camerer *et al.* (2018) demonstrated, above, even papers in prestigious journals can fail to replicate. Based on the rest of psychology we can perhaps anticipate that perhaps 30-40% or more of widely cited studies in our field would not replicate.

It is unlikely that many of the classic studies will be the subject of replication studies. However, it may be possible to judge already published papers as to the likelihood that the results would be likely to be replicable. This is suggested by a second part of the study by Camerer *et al.* (2018) (Section §5a) into the reproducibility of social science studies published in *Nature* and *Science*. In parallel to the replication of the previously published experiments, a panel of several hundred volunteers from the scientific community was recruited to evaluate the already published version of the papers. In surveys and a kind of 'prediction market', the peer volunteers, reading the paper and the plans for replication, were able to predict (with impressive accuracy) how likely a paper would be to replicate (before the replication was carried out). Why didn't the original

reviewers and editors detect that some studies were unlikely to replicate, before passing the studies for publication. Was it that they, like all of us, were seduced by the exciting, interesting story that they wanted to believe?

Perhaps we should be carrying out a similar exercise to establish which of our classic papers are likely to be replicable and preferentially citing those papers. Even better we should be thinking how best to use replications to firm up our knowledge.

Replicating studies is hard work and can take longer than the original study, especially when the power of the experiments is increased by having many more subjects involved. For early career researchers (ERCs) especially (Allen & Mehler 2019), there is the problem that many journals will not publish replications, successful or unsuccessful. In a pioneering policy, the Psychology and Cognitive Neuroscience section of *Royal Society Open Science* guarantees to publish close replications of any article published in the journal (the ‘Pottery Barn’ principle, ‘you broke it, you sort it’) and of articles from most other major journals too. Let’s hope that other journals will follow suit, especially when the replication is of a study previously published in that journal (‘prestigious’ journals are notoriously reluctant to do this). Funders also need to provide grants which include funding for replications.

Even when later studies show that a result was probably a false-positive, citations to debunked papers or concepts continue undiminished. This contributes to the long half-life of erroneous science. You will know your own examples from semiochemical research. In the life sciences, we’re not self-correcting in the ways that science aspires to be (Chambers 2017, p 50 ff; Ioannidis 2012). A key problem is that we currently have no system in biology for systematic forward-going flags linking papers and their replications (failed or successful) (discussed in e.g. Huber *et al.* 2019). Instead, when citing the earlier literature each of us should be searching forward on every occasion – which of course we don’t have time to do.

9. Human semiochemical targets for future investigation

The symposium (this issue) explored many potential human interactions that may be influenced by semiochemicals, either cues (probably the majority) or pheromones.

Where we don’t have candidate molecules, we should use the natural odours/secretions under investigation. There are many examples of scientists using this approach, as in the investigations of possible fear-associated odours (de Groot this volume; Pause this volume). These odours are likely to be cues, best understood at this stage as a response from the receiver but without an evolved emission of the odour by the ‘sender’. The responses to these cues may or not involve learning, either of context dependent cues, or, in other phenomena of recognition, individual odour signatures. Some phenomena may turn out to be pheromones: the first one to be confirmed may be human mother areola secretions (Schaal & Al Aïn 2014; Schaal this volume). Some may even be linked in the future with a probable receptor. Any future proposed pheromones will need to satisfy the steps outlined in Section §2b.

We should stop using androstadienone and estratetraenol as stimuli since there is no evidence that these are ‘human pheromones’, despite their wide use experimentally (see Section §3, (Doty 2014; Wyatt 2015; Wysocki & Preti 2004). There is a real opportunity cost: every experiment using these molecules is scientist time or funding not going to answer real questions about human semiochemistry.

New techniques offer ways of analysing the volatile odours given off by humans in different contexts in real time (Roberts this volume; Williams this volume). This will potentially free researchers from the limitations of having to take samples from particular parts of the body over a period of time, typically collecting odours on cotton pads under the armpits for example, which averages the collection over many hours.

What kinds of molecules should we be looking for? Since humans do not have a VNO, with V2Rs, the molecules are not likely to be large molecules, even if delivered at short range or contact (Wyatt 2015). It is unlikely therefore that we have protein pheromones like mouse darcin or ESP1. We don’t yet know if TAARs or other receptors in smaller numbers in MOE are important. It is as likely that small molecules are detected by regular ORs in the MOE.

While human semiochemicals are not likely to be large molecules, there has often been an unacknowledged assumption that pheromones, if we have them, would act at a distance of many body lengths, like female moth pheromones. Instead, the crucial interactions with an important role for semiochemicals (including as yet undemonstrated pheromones) in human sexual behaviour could be at very close range, in more intimate interactions (Wyatt 2015).

However, a key problem for researchers remains the subtlety and variety of human responses and the difficulty of designing semiochemical bioassays that are both biologically meaningful and repeatable. There is surely much to discover.

10. Acknowledgements

Thanks to [], and the anonymous referees for their helpful comments. All responsibility is mine.

11. References

- Allen, C & Mehler, DMA (2019) Open science challenges, benefits and tips in early career and beyond. *PLoS Biology* 17: e3000246.
- Bargh, JA, Chen, M & Burrows, L (1996) Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology* 71: 230.
- Begley, CG & Ellis, LM (2012) Raise standards for preclinical cancer research. *Nature* 483: 531.
- Bishop, DV & Thompson, PA (2016) Problems in using p-curve analysis and text-mining to detect rate of p-hacking and evidential value. *PeerJ* 4: e1715.
- Bishop, DV (2019) Rein in the four horsemen of irreproducibility. *Nature* 568: 435-435.

- Button, KS, Ioannidis, JP, Mokrysz, C, Nosek, BA, Flint, J, Robinson, ES & Munafo, MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews in Neuroscience* 14: 365-376.
- 575 Camerer, CF, et al (2018) Evaluating the replicability of social science experiments in *Nature and Science* between 2010 and 2015. *Nature Human Behaviour* 2: 637-644.
- Carney, DR, Cuddy, AJ & Yap, AJ (2010) Power posing: brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science* 21: 1363-1368.
- Carney, DR, Cuddy, AJ & Yap, AJ (2015) Review and Summary of Research on the Embodied Effects of Expansive (vs. Contractive) Nonverbal Displays. *Psychological Science* 26: 580 657-663.
- Carney, DR 2016 *My position on "Power Poses"* [Online]. Available: https://web.archive.org/web/20190723070448/http://faculty.haas.berkeley.edu/dana_carney/pdf/My%20position%20on%20power%20poses.pdf [Accessed 23 July 2019].
- 585 Cesario, J, Jonas, KJ & Carney, DR (2017) CRSP special issue on power poses: what was the point and what did we learn? *Comprehensive Results in Social Psychology* 2: 1-5.
- Chambers, C (2017) The seven deadly sins of psychology : a manifesto for reforming the culture of scientific practice. Princeton, NJ: Princeton University Press.
- Chambers, CD & Mellor, DT (2018) Protocol transparency is vital for registered reports. 590 *Nature Human Behaviour* 2: 791-792.
- Chambers, CD (2019a) The registered reports revolution Lessons in cultural reform. *Significance* 16: 23-27.
- Chambers, CD (2019b) The battle for reproducibility over storytelling. *Cortex* 113: A1-A2.
- Claesen, A, Gomes, SLBT & Tuerlinckx, F (2019) Preregistration: Comparing Dream to 595 Reality. *PsyArXiv*.
- Cuddy, AJC, Schultz, SJ & Fosse, NE (2018) P-Curving a More Comprehensive Body of Research on Postural Feedback Reveals Clear Evidential Value for Power-Posing Effects: Reply to Simmons and Simonsohn (2017). *Psychological Science* 29: 656-666.
- 600 Curran, T (2018) Reproducibility of academic preclinical translational research: lessons from the development of Hedgehog pathway inhibitors to treat cancer. *Open Biol* 8.
- de Groot, JH, Smeets, MA, Kaldewaij, A, Duijndam, MJ & Semin, GR (2012) Chemosignals communicate human emotions. *Psychological Science* 23: 1417-1424.
- de Groot, JH (this volume) Three degrees of fear: the chemical communication of emotion 605 intensity. *Phil Trans Roy Soc B*.
- Doty, RL (1981) Olfactory communication in humans. *Chemical Senses* 6: 351-376.
- Doty, RL (2014) Human pheromones: Do they exist? In: Mucignat-Caretta, C (ed.) *Neurobiology of chemical communication*. pp. 535-560. Boca Raton, FL: CRC Press.
- Doty, RL (ed.) (2015) *Handbook of olfaction and gustation. 3rd Edition*. Chichester: Wiley- 610 Blackwell.
- Doyen, S, Klein, O, Pichon, C-L & Cleeremans, A (2012) Behavioral priming: It's all in the mind but whose mind? *PLoS ONE* 7.

- Drea, CM, Boulet, M, Delbarco-Trillo, J, Greene, LK, Sacha, CR, Goodwin, TE & Dubay, GR (2013) The "secret" in secretions: methodological considerations in deciphering primate olfactory communication. *American Journal of Primatology* 75: 621-642.
- Drea, CM (2015) D'scent of man: A comparative survey of primate chemosignaling in relation to sex. *Hormones and Behavior* 68: 117-133.
- Ferdenzi, C, Delplanque, S, Atanassova, R & Sander, D (2016) Androstadienone's influence on the perception of facial and vocal attractiveness is not sex specific. *Psychoneuroendocrinology* 66: 166-175.
- Flanagan, KA, Webb, W & Stowers, L (2011) Analysis of male pheromones that accelerate female reproductive organ development. *PLoS ONE* 6: e16660.
- Frumin, I & Sobel, N (2013) An assay for human chemosignals. In: Touhara, K (ed.) *Pheromone Signaling*. pp. 373-394. New York: Humana Press.
- Gorgolewski, KJ & Poldrack, RA (2016) A Practical Guide for Improving Transparency and Reproducibility in Neuroimaging Research. *PLoS Biology* 14: e1002506.
- Hardwicke, TE & Ioannidis, JPA (2018) Mapping the universe of registered reports. *Nature Human Behaviour* 2: 793-796.
- Hare, RM, Schlatter, S, Rhodes, G & Simmons, LW (2017) Putative sex-specific human pheromones do not affect gender perception, attractiveness ratings or unfaithfulness judgements of opposite sex faces. *Royal Society Open Science* 4: 160831.
- Haselton, MG & Gildersleeve, K (2016) Human ovulation cues. *Current Opinion in Psychology* 7: 120-125.
- Havlicek, J (this volume) Olfaction, MHC and mate choice. *Phil Trans Roy Soc B*.
- Havlíček, J, Dvůřáková, R, Bartoš, L & Flegr, J (2006) Non advertized does not mean concealed: body odour changes across the human menstrual cycle. *Ethology* 112: 81-90.
- Havlíček, J, Murray, AK, Saxton, TK & Roberts, SC (2010) Current issues in the study of androstenes in human chemosignaling. In: Gerald, L (ed.) *Pheromones*. pp. 47-81. London: Academic Press.
- Huber, DE, Potter, KW & Huszar, LD (2019) Less "story" and more "reliability" in cognitive neuroscience. *Cortex* 113: 347-349.
- Hummer, TA & McClintock, MK (2009) Putative human pheromone androstadienone attunes the mind specifically to emotional information. *Hormones and Behavior* 55: 548-559.
- Huoviala, P & Rantala, MJ (2013) A Putative Human Pheromone, Androstadienone, Increases Cooperation between Men. *PLoS ONE* 8: e62499.
- Hurst, JL & Beynon, RJ (2013) Rodent urinary proteins used in scent communication. In: East, ML & Dehnhard, M (eds.) *Chemical signals in vertebrates* 12. pp. 117-133. New York: Springer.
- Ioannidis, JP (2010) Meta-research: The art of getting it wrong. *Res Synth Methods* 1: 169-184.

- Ioannidis, JP (2012) Why science is not necessarily self-correcting. *Perspectives on Psychological Science* 7: 645-654.
- Ishii, KK & Touhara, K (2018) Neural circuits regulating sexual behaviors via the olfactory system in mice. *Neuroscience Research*.
- Jacob, S & McClintock, MK (2000) Psychological state and mood effects of steroidal chemosignals in women and men. *Hormones and Behavior* 37: 57-78.
- 660 Jezierski, T, Dziecioł, M, Szumny, A, Nizański, W, Woszczyło, M, Pieczewska, B & Godzińska, EJ (2019) Discrimination of estrus odor in urine by male dogs in different experimental settings. *Journal of Veterinary Behavior* 29: 25-30.
- Jonas, KJ, Cesario, J, Alger, M, Bailey, AH, Bombari, D, Carney, D, Dovidio, JF, Duffy, S, Harder, JA, van Huistee, D, Jackson, B, Johnson, DJ, Keller, VN, Klaschinski, L, LaBelle, O, 665 LaFrance, M, M. Latu, I, Morssinkhoff, M, Nault, K, Pardal, V, Pulfrey, C, Rohleder, N, Ronay, R, Richman, LS, Mast, MS, Schnabel, K, Schröder-Abé, M & Tybur, JM (2017) Power poses – where do we stand? *Comprehensive Results in Social Psychology* 2: 139-141.
- Jorre De St Jorre, T, Hawken, PAR & Martin, GB (2014) New understanding of an old phenomenon: uncontrolled factors and misconceptions that cast a shadow over studies of the ‘male effect’ on reproduction in small ruminants. *Turkish Journal of Veterinary and Animal Sciences* 38: 625-636.
- Keller, A & Vosshall, LB (2004) Human olfactory psychophysics. *Current Biology* 14: R875-R878.
- 675 Kerr, NL (1998) HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review* 2: 196-217.
- Kjeldmand, L, Salazar, LTH & Laska, M (2011) Olfactory sensitivity for sperm-attractant aromatic aldehydes: a comparative study in human subjects and spider monkeys. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* 197: 15-23.
- 680 Klein, RA, Vianello, M, Hasselman, F, Adams, BG, Adams Jr, RB, Alper, S, Aveyard, M, Axt, JR, Babalola, MT & Bahník, Š (2018) Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science* 1: 443-490.
- 685 Lakens, D, Hilgard, J & Staaks, J (2016) On the reproducibility of meta-analyses: six practical recommendations. *BMC Psychol* 4: 24.
- Lapid, H & Hummel, T (2013) Recording odor-evoked response potentials at the human olfactory epithelium. *Chemical Senses* 38: 3-17.
- Lobmaier, JS, Fischbacher, U, Probst, F, Wirthmuller, U & Knoch, D (2018) Accumulating evidence suggests that men do not find body odours of human leucocyte antigen-dissimilar women more attractive. *Proc Biol Sci* 285.
- 690 Logan, DW, Brunet, JL, Webb, WR, Cutforth, T, Ngai, J & Stowers, L (2012) Learned recognition of maternal signature odors mediates the first suckling episode in mice. *Current Biology* 22: 1998-2007.

- 695 Lundström, JN, Gordon, AR, Alden, EC, Boesveldt, S & Albrecht, J (2010) Methods for
building an inexpensive computer-controlled olfactometer for temporally-precise
experiments. *International Journal of Psychophysiology* 78: 179-189.
- McGann, JP (2017) Poor human olfaction is a 19th-century myth. *Science* 356.
- Miyazawa, T, Gallagher, M, Preti, G & Wise, PM (2009) Methodological factors in odor
700 detection by humans. *Chemosensory Perception* 2: 195-202.
- Monti-Bloch, L & Grosser, BI (1991) Effect of putative pheromones on the electrical activity
of the human vomeronasal organ and olfactory epithelium. *Journal of Steroid
Biochemistry and Molecular Biology* 39: 573-582.
- Munafò, MR (2017) Improving the Efficiency of Grant and Journal Peer Review: Registered
705 Reports Funding. *Nicotine & Tobacco Research* 19: 773-773.
- Munafò, MR, Nosek, BA, Bishop, DVM, Button, KS, Chambers, CD, Percie du Sert, N,
Simonsohn, U, Wagenmakers, E-J, Ware, JJ & Ioannidis, JPA (2017) A manifesto for
reproducible science. *Nature Human Behaviour* 1: 0021.
- 710 Murata, K, Tamogami, S, Itou, M, Ohkubo, Y, Wakabayashi, Y, Watanabe, H, Okamura, H,
Takeuchi, Y & Mori, Y (2014) Identification of an olfactory signal molecule that
activates the central regulator of reproduction in goats. *Current Biology* 24: 681-686.
- Natsch, A (this volume) The biochemistry of human body odour: chemicals, bacteria and
enzymes. *Phil Trans Roy Soc B*.
- 715 Nelson, LD, Simmons, J & Simonsohn, U (2018) Psychology's Renaissance. *Annual Review of
Psychology* 69: 511-534.
- Nosek, BA, Ebersole, CR, DeHaven, AC & Mellor, DT (2018) The preregistration revolution.
Proceedings of the National Academy of Sciences of the United States of America 115:
2600-2606.
- Nuzzo, R (2015) Fooling ourselves. *Nature* 526: 182-185.
- 720 Olsson, MJ (this volume) Cues of sickness in body odour. *Phil Trans Roy Soc B*.
- Open Science Collaboration (2015) Estimating the reproducibility of psychological science.
Science 349.
- Pause, B (this volume) Superior cortical processing of anxiety- and aggression related
chemosignals in women. *Phil Trans Roy Soc B*.
- 725 Poldrack, RA, Baker, CI, Durnez, J, Gorgolewski, KJ, Matthews, PM, Munafò, MR, Nichols, TE,
Poline, J-B, Vul, E & Yarkoni, T (2017) Scanning the horizon: towards transparent
and reproducible neuroimaging research. *Nature Reviews Neuroscience* 18: 115.
- Ranehill, E, Dreber, A, Johannesson, M, Leiberg, S, Sul, S & Weber, RA (2015) Assessing the
robustness of power posing: no effect on hormones and risk tolerance in a large
730 sample of men and women. *Psychological Science* 26: 653-656.
- Roberts, SC (this volume) Tracking context-dependent odour changes in real time. *Phil
Trans Roy Soc B*.
- Schaal, B & Porter, RH (1991) Microsmatic humans revisited - the generation and
perception of chemical signals. *Advances in the Study of Behavior* 20: 135-199.

- 735 Schaal, B, Coureaud, G, Langlois, D, Ginies, C, Semon, E & Perrier, G (2003) Chemical and behavioural characterization of the rabbit mammary pheromone. *Nature* 424: 68-72.
- Schaal, B & Al Aïn, S (2014) Chemical signals ‘selected for’ newborns in mammals. *Animal Behaviour* 97: 289-299.
- 740 Schaal, B (this volume) Olfaction in the engagement and development of adaptive reciprocity in human infants and parents. *Phil Trans Roy Soc B*.
- Schwende, FJ, Wiesler, D, Jorgenson, JW, Carmack, M & Novotny, M (1986) Urinary volatile constituents of the house mouse, *Mus musculus*, and their endocrine dependency. *Journal of Chemical Ecology* 12: 277-296.
- 745 Shirasu, M & Touhara, K (2011) The scent of disease: volatile organic compounds of the human body related to disease and disorder. *Journal of Biochemistry* 150: 257-266.
- Simmons, J, Nelson, L & Simonsohn, U. 2018a. [66] Outliers: Evaluating A New P-Curve Of Power Poses. *DataColada.org* [Online]. Available from: <http://datacolada.org/66> [Accessed 03 August 2019].
- 750 Simmons, JP, Nelson, LD & Simonsohn, U (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22: 1359-1366.
- Simmons, JP & Simonsohn, U (2017) Power Posing: P-Curving the Evidence. *Psychological Science* 28: 687-693.
- 755 Simmons, JP, Nelson, LD & Simonsohn, U (2018b) False-Positive Citations. *Perspect Psychol Sci* 13: 255-259.
- Simonsohn, U, Nelson, LD & Simmons, JP (2014) P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General* 143.
- Simonsohn, U, Nelson, LD & Simmons, JP (2019) P-curve won't do your laundry, but it will distinguish replicable from non-replicable findings in observational research: Comment on Bruns & Ioannidis (2016). *PLoS ONE* 14: e0213454.
- 760 Smith, TD, Laitman, JT & Bhatnagar, KP (2014) The shrinking anthropoid nose, the human vomeronasal organ, and the language of anatomical reduction. *The Anatomical Record* 297: 2196-2204.
- 765 Sorensen, PW & Wisenden, BD (eds.) (2015) *Fish pheromones and related cues*. Hoboken: John Wiley.
- Vazire, S (2018) Implications of the Credibility Revolution for Productivity, Creativity, and Progress. *Perspect Psychol Sci* 13: 411-417.
- Wang, D, Forstmeier, W, Ihle, M, Khadraoui, M, Jerónimo, S, Martin, K & Kempnaers, B
- 770 (2018) Irreproducible text-book “knowledge”: The effects of color bands on zebra finch fitness. *Evolution* 72: 961-976.
- Wilke, K, Martin, A, Terstegen, L & Biel, S (2007) A short history of sweat gland biology. *International Journal of Cosmetic Science* 29: 169-180.
- Williams, J (this volume) What can we learn from air chemistry measurements of crowds?
- 775 *Phil Trans Roy Soc B*.
- Wyatt, TD (2009) Fifty years of pheromones. *Nature* 457: 262-263.

- Wyatt, TD (2010) Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and vertebrates. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* 196: 685-700.
- Wyatt, TD (2014) Pheromones and animal behavior: Chemical signals and signatures, 2 edn. Cambridge: Cambridge University Press.
- Wyatt, TD (2015) The search for human pheromones: the lost decades and the necessity of returning to first principles. *Proceedings of the Royal Society of London B: Biological Sciences* 282: 2014.2994.
- Wyatt, TD (2017) Pheromones. *Current Biology* 27: R739–R743.
- Wysocki, CJ & Preti, G (2004) Facts, fallacies, fears, and frustrations with human pheromones. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology* 281A: 1201-1211.
- Yong, E (2012) Replication studies: Bad copy. *Nature* 485: 298-300.
- Yong, E. 2018 Psychology's replication crisis is running out of excuses. *The Atlantic* [Online]. Available: <https://www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223/> [Accessed 15 August 2019].
- Zhou, W & Chen, D (2008) Encoding human sexual chemosensory cues in the orbitofrontal and fusiform cortices. *Journal of Neuroscience* 28: 14416-14421.