



More Data, Please: Machine Learning to Advance the Multidisciplinary Science of Human Sociochemistry

Jasper H. B. de Groot^{1,2*}, Ilja Croijmans¹ and Monique A. M. Smeets¹

¹ Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, Netherlands, ² Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

OPEN ACCESS

Edited by:

Marianne Schmid Mast,
University of Lausanne, Switzerland

Reviewed by:

Vivian P. Ta,
Lake Forest College, United States
Laiquan Zou,
Southern Medical University, China

*Correspondence:

Jasper H. B. de Groot
J.deGroot@psych.ru.nl

Specialty section:

This article was submitted to
Personality and Social Psychology,
a section of the journal
Frontiers in Psychology

Received: 09 July 2020

Accepted: 05 October 2020

Published: 22 October 2020

Citation:

de Groot JHB, Croijmans I and
Smeets MAM (2020) More Data,
Please: Machine Learning to Advance
the Multidisciplinary Science
of Human Sociochemistry.
Front. Psychol. 11:581701.
doi: 10.3389/fpsyg.2020.581701

Communication constitutes the core of human life. A large portion of our everyday social interactions is non-verbal. Of the sensory modalities we use for non-verbal communication, olfaction (i.e., the sense of smell) is often considered the most enigmatic medium. Outside of our awareness, smells provide information about our identity, emotions, gender, mate compatibility, illness, and potentially more. Yet, body odors are astonishingly complex, with their composition being influenced by various factors. Is there a chemical basis of olfactory communication? Can we identify molecules predictive of psychological states and traits? We propose that answering these questions requires integrating two disciplines: psychology and chemistry. This new field, coined *sociochemistry*, faces new challenges emerging from the sheer amount of factors causing variability in *chemical composition* of body odorants on the one hand (e.g., diet, hygiene, skin bacteria, hormones, genes), and variability in *psychological* states and traits on the other (e.g., genes, culture, hormones, internal state, context). In past research, the reality of these high-dimensional data has been reduced in an attempt to isolate unidimensional factors in small, homogenous samples under tightly controlled settings. Here, we propose big data approaches to establish novel links between chemical and psychological data on a large scale from heterogeneous samples in ecologically valid settings. This approach would increase our grip on the way chemical signals non-verbally and subconsciously affect our social lives across contexts.

Keywords: sense of smell, machine learning, chemosignals, non-verbal communication, social and personality psychology

INTRODUCTION

Humans are surprisingly good smellers. The pervasive myth that humans are only “tiny smellers” has been debunked by 21st century research showing a wide array of smell skills (Stevenson, 2010; de Groot et al., 2017; McGann, 2017). To name a few: humans can follow a scent-trail (like sniffer dogs; Porter et al., 2007), detect certain odorants at extremely low levels (few droplets in an Olympic size swimming pool; Yeshurun and Sobel, 2010), and identify diseases like Parkinson’s before actual diagnosis (Trivedi et al., 2019). In our everyday lives, smells have a “communicative” function, informing us about the quality of food and warning us for environmental hazards (e.g., gas leaks) (Stevenson, 2010). An even less well-known function of smell is

social communication (de Groot et al., 2017; Parma et al., 2017; Pause, 2017; Roberts et al., 2020); the topic of this article. Studies have shown that our smells provide others with cues about our identity and gender (Penn et al., 2007), age (Mitro et al., 2012), health (Olsson et al., 2014), and emotions (de Groot et al., 2015; Pause et al., 2020). This form of communication occurs without our voluntary control and generally outside of our awareness, which imbues chemical communication with mystery. Demystifying the spreading of social information through smell was listed in *Science* as one of the 125 most compelling multidisciplinary puzzles facing scientists this century (Kennedy and Norman, 2005). Our goal here is to outline *how* researchers could go about answering this query, whether there is a universal “language” of social smells. Society at large will be helped by optimally leveraging fundamental insights emerging from this view to worldwide industrial and clinical applications that could improve a person’s quality of life.

Social smells are markedly complex: body odor contains thousands of molecules (de Lacy Costello et al., 2014), and massive variability is caused by factors including genotype, hormonal status, mood, skin bacteria, diet, smoking, hygiene habits, clothing, and use of fragranced products (e.g., Natsch and Emter, 2020; Roberts et al., 2020). Past studies have generally sidestepped this challenge by performing small-scale psychological experiments under carefully controlled, sterile conditions (for a meta-analysis: de Groot and Smeets, 2017; for a critical view: Wyatt, 2020). These studies formed the first stepping stones by strongly suggesting that social information can be communicated via smell under tightly controlled settings; yet, (i) the molecules transmitting the message have generally remained elusive, as well as (ii) the ecological settings in which chemical communication occurs. Not dealing with these obstacles could deadlock future research efforts to “anchor” molecules to their social source. To accelerate future research, we propose (i) multidisciplinary ways of working by integrating psychology and chemistry toward a science of human *sociochemistry* (Box 1), and (ii) moving outside of the sterile lab to test subjects with diverse backgrounds.

The sociochemistry we advance is a multidisciplinary, ecological approach that in view of its inherent complexity requires an ecosystem of academic institutions around the world to flourish, by working together to create speed and scale (cf. Forscher et al., 2020). We propose the building of open access databases holding information that spans across chemistry (e.g., chemical composition of sweat odor) and psychology (e.g., capturing the states and traits of those participating in the chemosignaling as well as their unique contextual information. Machine learning techniques can be applied to generate models that may accurately predict molecules’ sway on our social lives across diverse contexts and samples, with technological, societal, and clinical applications following suit.

BOX 1 | Definition of sociochemistry. With sociochemistry we refer to the multidisciplinary science examining non-verbal social communication via human body odor, particularly focusing on the chemistry between people.

In what follows, we will first outline the initial research questions, methods, and advances of past research, before we will identify current obstacles to a broader and deeper understanding of human chemical communication, and we will end with a perspective on how to overcome these hurdles in future research.

PAST RESEARCH: SIMPLIFYING A COMPLEX PROBLEM

Communication is crucial to humans. Most of our communication is non-verbal. Of all the sensory channels engaged in non-verbal communication, smells arguably pose the biggest deciphering challenge.

The past research in this field initially focused on determining *what* social information can be communicated via smell. To test this, researchers have systematically attempted to eradicate “noise” on the chemical communication channel by controlling extraneous factors (e.g., diet, hygiene, fragranced product use) and testing homogeneous samples in carefully controlled lab experiments. In these studies, sweat was collected from *senders* (who had kept to a scent-free regimen for multiple days to isolate the experimentally-induced chemical “message”) and presented to *receivers* in a separate experiment. Chemical communication was inferred from recipients’ behavioral, affective, physiological, neuroendocrine and/or neural responses matching the sender’s state. This way, numerous double-blind experiments showed that human smells can convey information from fleeting emotions and sickness, to more enduring traits like identity, gender, reproductive status, and age (for reviews, see de Groot et al., 2017; Parma et al., 2017; Pause, 2017).

Although past research on human chemical communication has provided initial insights into the type of information human odors can bring across, we identify a number of obstacles for a better (quicker, broader, and deeper) understanding of non-verbal communication via smell.

Problem I: Small Scale, Slow Speed

The current science of non-verbal communication via smell is rooted in a longstanding tradition of strictly controlled laboratory experiments focusing on the empirical testing of hypotheses addressing cause-effect relations, using reliable and validated methods and carefully calibrated instruments (for empirical demonstrations, see e.g., Chen and Haviland-Jones, 2000; Regenbogen et al., 2017; Endevelt-Shapira et al., 2018; Quintana et al., 2019; de Groot et al., 2020b; Gomes et al., 2020; Pause et al., 2020 (for recent narrative overviews, see e.g., Loos et al., 2019; Ferdenzi et al., 2020; Havlíček et al., 2020 (for meta-analyses, see e.g., Gildersleeve et al., 2014; de Groot and Smeets, 2017). This approach, with a preference for intrinsic over extrinsic validity, has been the method of choice to build our (psychological) science for decades. Despite advantages of scientific rigor and quality, there are problems in speed and scale. With little coordination across labs around the world, different researchers may be working on similar research questions (e.g., “can humans smell fear?”; Mujica-Parodi et al., 2009; Pause et al., 2009; Prehn-Kristensen et al., 2009; Zhou and Chen, 2009), each

moving through the laborious cycle of recruiting, screening, and testing senders and receivers with barely sufficient statistical power (as outlined by Wyatt, 2020). The essence to our argument is that there is a stark contrast between the *complexity* of the problem at the root of sociochemistry, which is the mystery of the correspondence between the chemical “code” and the message it carries on the one hand, and the relatively slow tactic of churning experiments one at a time.

Problem II: Generality of Findings

Second, we need to characterize the generality of findings or extrinsic validity of the traditional experiments (Simons et al., 2017). Both uniformity in subject characteristics and test settings form obstacles to a broader understanding of the potentially species-wide and real-world impact of non-verbal communication via smell. Open queries include: Is the language of smell universal? How much of this communication is modified by context, a powerful moderating factor in olfactory science (e.g., Dalton, 1999; De Araujo et al., 2005; de Groot et al., 2020a)? Can this language be “heard” beyond the thick walls of labs, in noisy field settings? Answering these questions will help chart the impact of social smells on the daily lives of many.

Because past research has been typified by (i) context-deprived lab experiments, presenting (ii) uncontaminated sweat samples, using (iii) a relatively small number of subjects with (iv) relatively uniform characteristics, we currently have no knowledge of how broadly shared human olfactory communication is. To illustrate, the male-to-female chemical communication dyad initially served to increase experimental sensitivity, with males generally having the larger and more active sweat glands, and females being the slightly better smellers (but see this meta-analysis: Sorokowski et al., 2019; and this review: Majid et al., 2017, for gender differences that are at most small and affecting only higher order smell processing). Although initially useful, this gender uniformity adds a constraint on generality, and the same goes for the almost exclusive reliance on participants that are Western, Educated, Industrialized, Rich, Democratic (WEIRD; Henrich et al., 2010) (cf. de Groot et al., 2018; Roberts et al., 2020). Generalizing research findings from WEIRD samples to other populations is a major problem in science in general, and a particularly pressing issue when one examines the breadth and scale of the non-verbal language of smells (Box 2).

Problem III: Unidisciplinary Research

Third, to be able to forge a link between smell molecules and behavior we need to move beyond a single-discipline research

BOX 2 | Sociochemical language. A sociochemical language would imply configurations of chemical symbols that convey meaning, which meaning is acquired via learning. This notion of language would acknowledge the possibility that (i) identical chemical configurations do not mean the same to everyone, (ii) the meaning of an identical chemical configuration may vary even to a single individual depending on context, (iii) that there is (substantial) variation or “noise” around one chemical configuration, from which one single uniform meaning can still be distilled. Therefore, the language would not have to be universal.

tradition. Although several psychological studies have revealed systematic patterns in the behaviors of senders and recipients (in relatively sterile, uniform settings), the *chemical message* driving this coupling has generally remained enciphered (but see Penn et al., 2007; Smeets et al., 2020). Lessons can be learned from the animal literature, where the combination of rigorous behavioral experiments (bioassay) and chemical analysis (isolating, identifying, and synthesizing the bioactive substance to recreate the bioassay-behavior) forms the golden standard to detect a common chemical “language” for a species: *pheromones* (Wyatt, 2015, 2020). But the definition of pheromones, rooted in entomological research as single molecules eliciting innate responses in a conspecific (Karlson and Lüscher, 1959), appears outdated and unsuitable for mammals like humans, as our smell perception strongly depends on learning and context, and our body emits a multitude of molecules (de Groot et al., 2017). The minimum pragmatic evidence, however, is to determine (in a collaborative, multi-lab effort) whether human chemical language is *consistent in form* (requiring a multidisciplinary approach) and *broadly shared* across the human species (requiring diverse samples and settings).

PROSPECTIVE ADVANCES

In the wake of recent developments in psychological research and theory, chemical analytical technology, and data science (discussed below), substantial progress can be made now to unravel the symbol system of social smell. Specifically, we outline an integration of traditional psychology methods and chemistry toward a new science of human sociochemistry, studying human chemosignaling across various ecologically valid settings and samples, across all human diversity. To deal with the complexity and large, multidimensional databases that emerge from this interdisciplinary, ecologically valid endeavor, we propose applying data science approaches like machine learning. We anticipate that large scale multidisciplinary collaborations are required to get us closer to identifying the alphabet of the language of social smells and assess its real-world impact.

Multidisciplinary Approach: Deciphering the Alphabet of Social Smells

Any attempt to get closer to the answer of whether social smells convey a common language requires a multidisciplinary combination of psychological experiments and chemical analysis.

Most research on human chemical communication focused on psychological effects. The few studies that did apply chemical analysis have shown that certain characteristics and transient emotions could be identified in a sender’s body odor. One pioneering study by Penn et al. (2007) showed that a person’s identity and gender could be expressed in a person’s body odor, with 14 molecules predicting gender with 75% accuracy. Based on remarkable anecdotal evidence that a human “super smeller” could detect Parkinson’s Disease (PD) by smell, Trivedi et al. (2019) found that four compounds (eicosane, hippuric acid, octadecanal, and perillaldehyde) were characteristic markers

of PD; when smelling these compounds, the super smeller subjectively reported a strong PD smell. Other studies found chemical markers suggestive of fear (and happiness). Potential chemical markers for fear were identified by in armpit odor (Smeets et al., 2020), stress levels were also expressed in a person's breath (Preti et al., 2019; acetone, isoprene, dimethyl sulfide), and in a creative field study, (Williams et al., 2016) showed that scary and funny film events reliably changed the emission of molecules from cinema audiences. Taken together, these multidisciplinary studies show the potential for social information to be encoded in a person's smell in predictable ways, thus jumpstarting a sociochemistry approach to identify a common smell language.

Whereas on the one end of the scientific spectrum, we have this classic tradition of sequentially conducting laboratory experiments designed to address a specific causal hypothesis derived from theory, carefully controlling for measurement error and extraneous influence. On the other end there is the big data approach relying on machine learning analytical techniques performed on big databases holding what seems to be unrelated information from large populations to magically reveal unexpected correlations unencumbered by theory (Mayer-Schönberger and Cukier, 2013). Neither, on its own, will be an optimal path for unraveling human sociochemistry and the underlying language on which it is built. What we propose, instead, is a hybrid approach, a combination in which machine learning techniques are used to help us find handles on and insights into the composition of the chemical signal combinations that are the building blocks of the signal, and the related individual and external variables to further sculp this unique form of social communication. These insights will contribute to the formulation of hypotheses about cause and effect that can then be isolated and tested in controlled lab environments (cf. Wyatt, 2015, 2020).

Ecological Validity: A Broadly Shared, Widely Used Social Smell Language?

In the quest for discovering a potential universal language of smell that is also societally relevant, we argue that the highest success rate can be achieved by first examining smells whose detection generally aids survival (Schaal and Porter, 1991).

In the earliest stages of life, when vision and hearing are still underdeveloped, the smell of mothers' milk is a powerful cue that attracts a newborn to the food source (Schaal et al., 2020). Even formula-fed newborns oriented more toward the smell of an unfamiliar lactating woman than to the familiar formula smell (Porter et al., 1991); and this was not a novelty effect, as the same smell was also preferred over the breast odor of nulliparous women (Makin and Porter, 1989). There may well be universal chemical cues in the breast odor of lactating women that attract most if not all newborns under diverse ecologically valid settings, but this still requires empirical investigation from non-WEIRD samples (Schaal et al., 2020).

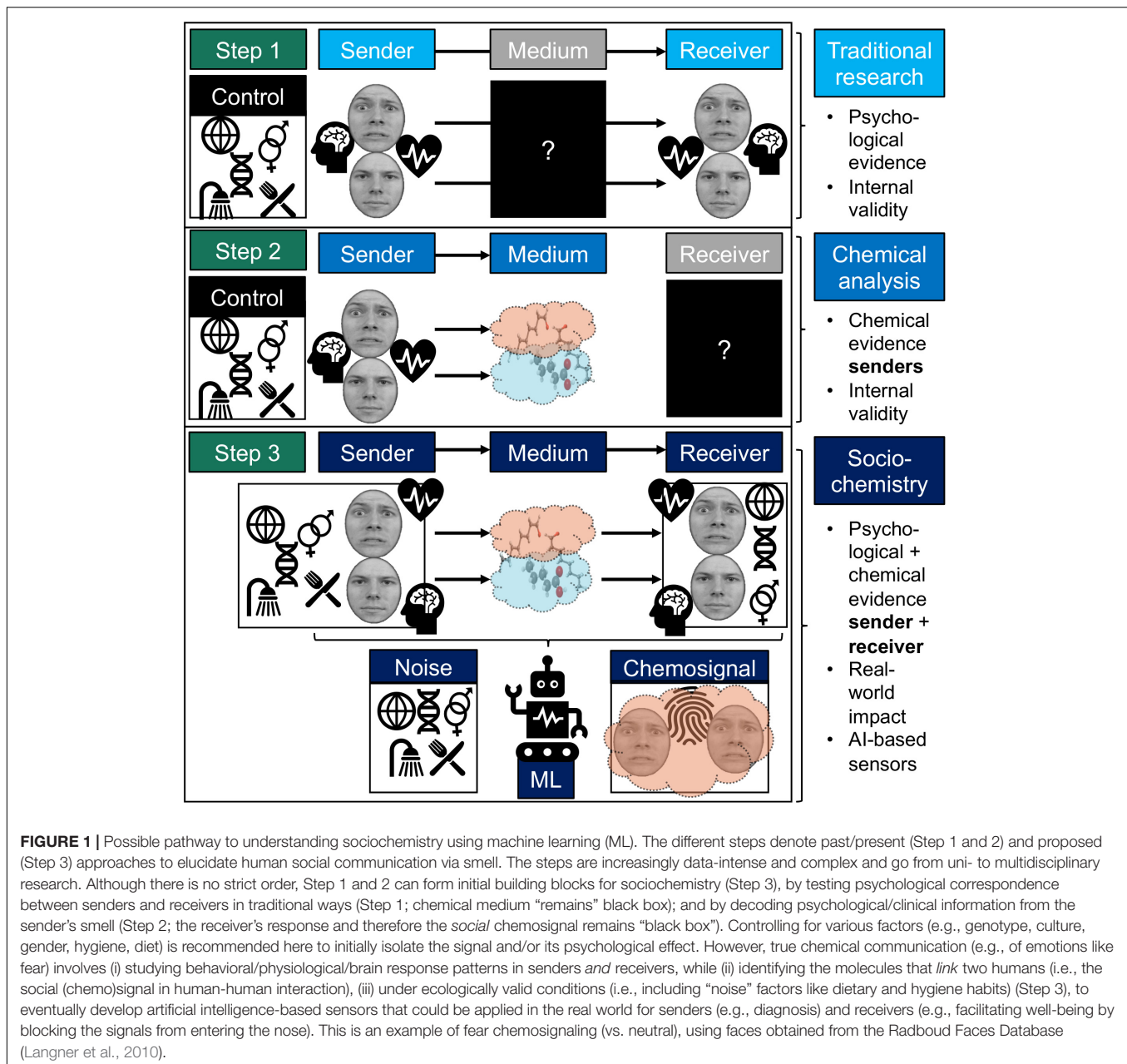
Humans would also benefit from picking up smells indicating danger, like fear sweat threatening physical harm, and disease sweat threatening contamination. The capacity to register

these invisible, far-reaching, and long-lasting chemical warning cues would have increased our ancestors' survival chances. Indeed, the smell of fear has been shown to instigate adaptive processes: a fearful facial expression (raised eyebrows, opened nose) and increased sensory intake (eyes and nose) to better detect threat (de Groot et al., 2012); yet, typically this phenomenon has not been examined beyond WEIRD samples, with one East Asian exception (de Groot et al., 2018). Quintana and colleagues (2019) further assessed the breadth of chemical communication in a controlled yet ecologically valid Virtual Reality environment. They found that smelling fear/stress sweat induced anxiety in recipients and reduced their interpersonal trust toward a virtual character. Even outside of the lab, the smell of fear (masked in clove odor, making it undetectable) could negatively impact dental student performance (Singh et al., 2018). Indeed, odor masking (e.g., with perfume, deodorant) could not prevent recipients from making consistent and reliable smell-based social judgments at typical social distances (Gaby and Zayas, 2017). Taken together, these findings allude to fear/stress smell affecting behavior across contexts in diverse samples, but more data is needed.

The complex and resource-intensive methodology of sweat sampling and exposure has arguably held back large (field) experiments, but upscaling and including natural settings seems inevitable in an attempt to discover the commonalities in human smells and their practical application, with big data approaches providing structure within the anticipated wealth of transdisciplinary data.

Machine Learning: Solving the Big Data Challenge Ahead of Us

In vision and hearing, the wavelength of light and frequency of sound are highly predictive of color and tone; yet, predicting the smell of a molecule from its chemical structure is much harder. In the past decade, researchers have started using machine learning techniques to demonstrate links between molecular structure and odor perception (for an overview: Lötsch et al., 2018). Machine learning, a popular application of artificial intelligence, is a set of methods that can be used to automatically detect patterns in data and use these patterns to predict or classify future data (e.g., Murphy, 2012; Dhar, 2013). Although machine learning models have shown the feasibility of predicting odor perception from relatively simple, non-social smells (Khan et al., 2007; Zarzo, 2011; Snitz et al., 2013; Keller et al., 2017; Gutiérrez et al., 2018; Sanchez-Lengeling et al., 2019) a number of extra challenges emerge when machine learning is applied to uncover the language of social smells. The difference between past "non-social" models and what we propose here is that (i) past models predicted odor perception from physico-chemical properties of *single* chemical compounds, whereas body odors are mixtures of compounds, and the communicative signal also likely having a multi-component architecture (Loos et al., 2014), the composition of which requires employing chemical analytical techniques to elucidate; (ii) past model endpoints have traditionally been



sensory endpoints (e.g., intensity, pleasantness, and qualitative descriptors like garlicky or fruity) as opposed to social-behavioral endpoints (e.g., perceivers’ affect, physiology, behavior); (iii), past models have not considered various sample characteristics (excepting gene variants coding for odorants receptors) or ecologically relevant contexts that are expected to impact smell perception as well.

To identify human chemosignals within the vast amount of data that can encompass body odors (a big data challenge), we recommend moving away from using a single, traditional statistical model (e.g., logistic regression), and instead propose a sequence of different analyses, including machine learning (ML). It would seem premature to rigorously define each step in the

analysis sequence, but we will sketch a possible analysis “pipeline” (Figure 1):

Step 1 would entail collecting sweat from senders induced to be in a particular state (e.g., fear, happiness, disgust, sickness) or having a characteristic of interest (e.g., gender, personality, genotype). A subset of these sweat samples would then be used as *stimuli* in another experiment involving human receivers, whose behavioral responses will form a benchmark for verifying effective chemical communication (requiring a sender and receiver).

In **Step 2**, the remaining sweat samples will be used for chemical analysis. After extracting the molecules using headspace, solvent, or direct extraction techniques, chemical analysis could entail two-dimensional gas chromatography-mass

spectrometry (GCxGC-ToF-MS) allowing for comprehensive profiling of the volatile molecules in the sweat samples and their discriminative power between two (or more) states/traits of interest. Because there is little to no background knowledge on chemical classes associated with presumed signals in sweat odor, initial research by Smeets et al. (2020) used *untargeted* screening approaches to distinguish between fear, happiness, and a neutral state, and found a matrix of over a 1,000 chemical volatile peaks. This number could be reduced as a next step to 94 by selecting only those peak intensities that differed significantly with at least one other emotion category. Preprocessing the GC × GC-ToF-MS profiles into total-intensity-count values (TICs) is another way to yield a smaller, more manageable subset of peaks of interest (cf. Lebanov et al., 2020). What could further ease the future identification of unique chemical profiles predicting human states/traits are *templates* (reference peak profiles) that follow from overlaying all chromatograms in a set (cf. Stilo et al., 2019), or using previous datasets as templates (cf. Reichenbach et al., 2019). This requires acquiring large chemical datasets, which necessitates high-throughput approaches like automated extraction and (ultra)fast GCxGC-ToF-MS, followed by automated quantification of specific target compounds belonging to specific states and traits.

In **Step 3**, ML techniques could help identify the core chemical features of human states/traits in multiple ways. Unsupervised learning (e.g., k-means clustering) could yield potentially interesting clusters of chemicals that are involved in chemical communication not considered before. Supervised learning could be applied next by training an algorithm on a large subset of samples, and testing the trained model on the remaining set. While there are vast varieties in learning algorithms, they can broadly be divided into linear or non-linear based on the shape of the decision surface used to classify data. Linear methods, like support vector machines (SVM) with linear kernels, may be preferred because they perform at least on par with non-linear methods (e.g., Misaki et al., 2010, in the context of separating emotions with fMRI data) while remaining straightforward to interpret. The interpretability of the models from the pipeline we propose might be tested by comparing the predictive power of those models with the outcomes on receiver experiments. To illustrate, Reichenbach et al. (2019) combined GC × GC chemical profiling with SVM to predict different characteristics of wines (e.g., grape variety, origin). Although the wines had considerable overlap in their chemical composition (up to 25% overlap in grape variety), the analysis yielded a number of highly distinctive molecules that the models used to differentiate the wines with around 90% accuracy (Reichenbach et al., 2019). At the same time, the resulting models were still relatively intuitively interpretable (cf. Mori and Uchihira, 2019).

We believe that applying a chemosignal-identification pipeline as described above would also yield relatively straightforward models, with an interpretable set of chemical predictors that are highly predictive of the emotions under investigation. Feature selection in our machine learning pipeline could be based on, depending on the ML technique used, mean absolute error (MAE) of the predictor (in case of regression-based

techniques) or area under the curve (AUC) measures (in case of classification-based techniques) (Molnar, 2018). Selection of the best performing predictors could be tested by application level evaluation (cf. Molnar, 2018), using follow-up lab or crowd-sourced experiments where the most likely molecule candidates are tested in appropriate molecular concentrations.

DATABASE-BUILDING: BACK TO THE FUTURE

The proposed analysis pipeline requires rather large, well-populated databases compared to current standards. At present, there is a lack of such (publicly available) databases. Ideally, data in these databases contain a vast amount of parameters from hundreds of participants (senders *and* receivers). These parameters include personal factors (e.g., gender, age, country of residence, genotype), lifestyle factors (e.g., deodorant use, hygiene habits), measures of context (e.g., sterile lab vs. field), health, personality, and emotion (e.g., subjectively reported emotions and psychophysiological measures), and thousands of additional parameters per sample resulting from chemical analysis. Hence, the complexity and vastness of the resulting database underscores the need to step away from experimenter-driven analyses techniques such as traditional regression models, and turn to automatic feature selecting analyzation algorithms instead. Using these ML techniques has another advantage – the possibility to directly apply the best performing models in artificial intelligence applications.

However, one big hurdle to take with this multivariate, machine learning approach is the need for large, ecologically valid datasets. Talking about big data, a now famous 1989 National Geographic Smell Survey managed to test and analyze data from 1.42 million respondents to examine the relation between olfaction and aging (Wysocki and Gilbert, 1989). In a related effort to build socially relevant smell databases, Snitz et al. (2019) cleverly combined online crowdsourcing with the physical distribution of “scratch-and-sniff” odorants (via regular mail), and collected data (now publicly available) from about 1,000 individuals in 100 days. In these studies, chemical communication had not been the focus. If body odors were the topic, the database should hold matrices of (i) the chemical constituents of body odors (alongside multivariate information about, e.g., the emotional state during which the body odor was produced), as well as (ii) relevant person- and situation-specific variables of senders donating the samples (e.g., diet, hygiene product use, culture, genetic variation), and (iii) person- and situation-specific variables of recipients.

Acquiring such a complex and elaborate database can impossibly be a single-lab endeavor, and in our view, coordination within a larger ecosystem of labs will be crucial. Fortunately, technical advances in communication allow large consortia of researchers to globally collaborate from the comfort of their homes, like the Global Consortium for Chemosensory Research (GCCR), which focuses on the relation between COVID-19 and chemosensory dysfunction. Within weeks, hundreds of researchers around the globe collaborated to design an online study resulting in a large (open access) database

on COVID-19 and smell/taste dysfunction and a published manuscript (Parma et al., 2020). Moreover, in work that focused on smell communication (literally, being able to talk about smells), Majid et al. (2018) have examined for 20 languages whether there exists a universal hierarchy to vision being more accessible to consciousness and linguistic description than smell. These, and other examples (e.g., Iravani et al., 2020), have illustrated that global consortia can be instrumental in acquiring the necessary datasets to solve complex and urgent questions in a timely manner.

TECHNOLOGY SUPPORTING SOCIETAL AND CLINICAL IMPACT

We envision the application of machine learning to understand human non-verbal communication to yield a series of impactful consequences ranging from psychology to medicine. If machine learning techniques can pick up on statistical regularities between, for instance, emotional states and health conditions on the one hand and patterns of molecules on the other, chemical sensors can be developed to read this “smell language” in real time. Promisingly, Imam and Cleland (2020) placed an array of 72 chemosensors (based on the architecture of the mammalian olfactory bulb) in a wind tunnel, which rapidly learned and identified odor representations, despite various sources of noise. Given that body odor (itself susceptible to noise) contains information about emotions (Smeets et al., 2020) and one’s health, ranging from markers for Parkinson’s disease (Trivedi et al., 2019), general inflammatory reactions (Olsson et al., 2014), to possibly the presence of COVID-19 in sweat (Grandjean et al., 2020), it would be intriguing to explore whether such algorithms could be used to learn and identify the even more complex language inherent to human odors.

The near future could see a rapid growth in the diagnostic implementation of sweat odor analysis that could happen outside of a lab or clinic in a person’s home, with the emergence of novel smartphone-based biosensors (Brasier and Eckstein, 2019). Through these smartphone-based on-skin biosensors, sweat compounds could become broadly available in databases as digital biomarkers. Such an in-home approach is expected to have a major influence on clinical and outpatient care, and could even prevent infectious diseases from spreading by suggesting self-quarantining. The impact of these biosensors may extend to therapeutic settings, where the smell-based detection of patients’ emotions (or lack thereof) could provide an insightful role in (online) therapeutic sessions. In sum, physicians and clinicians could foresee their instrumentation being expanded in the future by sensors and machine learning to more quickly, accurately, and safely get a grip on a disease or clinical problems and their prognosis (Chen et al., 2019).

CONCLUSION

Although scientific evidence has shown that the sense of smell serves a number of crucial functions in the daily life of humans,

including social communication (e.g., Stevenson, 2010; de Groot et al., 2017; McGann, 2017), the idea that humans are micro smellers has remained hardwired among scientists and laypeople. However, through smell, humans can (unwillingly) convey information about a person. These initial advances were generally obtained under the most sterile conditions, by single research groups from the perspective of a single discipline. Although initially fruitful, we caution that continuing this experimental tradition will stall scientific progress toward a broader, deeper, and quicker understanding of non-verbal communication via smell. In the quest for discovering the real-world impact of social smells in diverse samples across diverse settings, we focused on the importance of ecological testing conditions, multidisciplinary research, and open collaborations to populate high dimensional databases, with machine learning approaches “making sense” of the complicated statistical regularities between smell molecules and physical or psychological conditions (the science of sociochemistry). By informing us about food, danger, health, and hygiene, olfaction serves a crucial role in human life, and so much so, that losing our sense of smell dramatically reduces the quality of our life. Our invitation for a better fundamental and practical understanding of the language of human smells opens up a multitude of (technological) possibilities, including tailor-made or world-wide clinical and societal applications proportionate to the scale at which human odors non-verbally communicate information from a sender to a recipient, whether human or machine.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors were involved in the conceptualization of this research. JG drafted the outline of this manuscript, wrote the manuscript, and edited the manuscript. IC and MS co-wrote the manuscript and critically revised the outline of the manuscript and the manuscript itself.

FUNDING

This research was supported by a Veni Innovational Research grant (NWO-016.Veni.195.116) awarded to the JG.

ACKNOWLEDGMENTS

We thank Daniel Oberski for his contributions on machine learning approaches and Helene Loos for her contributions on chemical analytical techniques (section “Machine Learning: Solving the Big Data Challenge Ahead of Us”).

REFERENCES

- Brasier, N., and Eckstein, J. (2019). Sweat as a source of next-generation digital biomarkers. *Digit. Biomark.* 3, 155–165. doi: 10.1159/000504387
- Chen, D., and Haviland-Jones, J. (2000). Human olfactory communication of emotion. *Percept. Mot. Skills* 91, 771–781. doi: 10.2466/pms.2000.91.3.771
- Chen, Z., Chen, Z., Song, Z., Ye, W., and Fan, Z. (2019). Smart gas sensor arrays powered by artificial intelligence. *J. Semicond* 40:111601. doi: 10.1088/1674-4926/40/11/111601
- Dalton, P. (1999). Cognitive influences on health symptoms from acute chemical exposure. *Heal. Psychol.* 18, 579–590.
- De Araujo, I. E., Rolls, E. T., Velazco, M. I., Margot, C., and Cayeux, I. (2005). Cognitive modulation of olfactory processing. *Neuron* 46, 671–679. doi: 10.1016/j.neuron.2005.04.021
- de Groot, J. H. B., Beetsma, J. V., van Aerts, T. J., Le Berre, E., Gallagher, D., Aarts, H., et al. (2020a). From sterile labs to rich VR: immersive multisensory context critical for odors to induce motivated cleaning behavior. *Behav. Res. Methods* 52, 1657–1670. doi: 10.3758/s13428-019-01341-y
- de Groot, J. H. B., Kirk, P. A., and Gottfried, J. A. (2020b). Titrating the smell of fear: initial evidence for dose-invariant behavioral, physiological, and neural responses. *Psychol. Sci.* (in press).
- de Groot, J. H. B., Semin, G. R., and Smeets, M. A. M. (2017). On the communicative function of body odors: a theoretical integration and review. *Perspect. Psychol. Sci.* 12, 306–324. doi: 10.1177/1745691616676599
- de Groot, J. H. B., and Smeets, M. A. M. (2017). Human fear chemosignaling: evidence from a meta-analysis. *Chem. Senses* 42, 663–673. doi: 10.1093/chemse/bjx049
- de Groot, J. H. B., Smeets, M. A. M., Kaldewaij, A., Duijndam, M. J. A., and Semin, G. R. (2012). Chemosignals communicate human emotions. *Psychol. Sci.* 23, 1417–1424. doi: 10.1177/0956797612445317
- de Groot, J. H. B., Smeets, M. A. M., Rowson, M. J., Bulsing, P. J., Blonk, C. G., Wilkinson, J. E., et al. (2015). A sniff of happiness. *Psychol. Sci.* 26, 684–700. doi: 10.1177/0956797614566318
- de Groot, J. H. B., van Houtum, L. A. E. M., Gortemaker, I., Ye, Y., Chen, W., Zhou, W., et al. (2018). Beyond the west: chemosignaling of emotions transcends ethno-cultural boundaries. *Psychoneuroendocrinology* 98, 177–185. doi: 10.1016/j.psyneuen.2018.08.005
- de Lacy Costello, B., Amann, A., Al-Kateb, H., Flynn, C., Filipiak, W., Khalid, T., et al. (2014). A review of the volatiles from the healthy human body. *J. Breath Res.* 8:014001. doi: 10.1088/1752-7155/8/1/014001
- Dhar, V. (2013). Data science and prediction. *Commun. ACM* 56, 64–73. doi: 10.1145/2500499
- Endevelt-Shapira, Y., Perl, O., Ravia, A., Amir, D., Eisen, A., Bezalel, V., et al. (2018). Altered responses to social chemosignals in autism spectrum disorder. *Nat. Neurosci.* 21, 111–119. doi: 10.1038/s41593-017-0024-x
- Ferdenzi, C., Richard Ortegón, S., Delplanque, S., Baldovini, N., and Bensafi, M. (2020). Interdisciplinary challenges for elucidating human olfactory attractiveness. *Philos. T. R. Soc. B* 375:20190268. doi: 10.1098/rstb.2019.0268
- Forscher, P. S., Wagenmakers, E., Coles, N. A., Silan, M. A., and IJzerman, H. (2020). A Manifesto for Team Science. *psyArxiv*[Preprint]. doi: 10.31234/osf.io/2mdxh
- Gaby, J. M., and Zayas, V. (2017). Smelling is telling: human olfactory cues influence social judgments in semi-realistic interactions. *Chem. Senses* 42, 405–418. doi: 10.1093/chemse/bjx012
- Gildersleeve, K., Haselton, M. G., and Fales, M. R. (2014). Meta-analyses and p-curves support robust cycle shifts in women's mate preferences: reply to Wood and Carden (2014) and Harris. Pashler, and Mickes (2014). *Psychol. Bull.* 140, 1272–1280. doi: 10.1037/a0037714
- Gomes, N., Silva, F., and Semin, G. R. (2020). The lasting smell of emotions: the effects of reutilizing fear sweat samples. *Behav. Res. Meth* doi: 10.3758/s13428-020-01412-5
- Grandjean, D., Sarkis, R., Tourtier, J.-P., Julien-Lecocq, C., Benard, A., Roger, V., et al. (2020). Detection dogs as a help in the detection of COVID-19 – Can the dog alert on COVID-19 positive persons by sniffing axillary sweat samples? Proof-of-concept study. *bioRxiv*[Preprint]. doi: 10.1101/2020.06.03.132134
- Gutiérrez, E. D., Dhurandhar, A., Keller, A., Meyer, P., and Cecchi, G. A. (2018). Predicting natural language descriptions of mono-molecular odorants. *Nat. Commun.* 9:4979. doi: 10.1038/s41467-018-07439-9
- Havlíček, J., Winternitz, J., and Roberts, S. C. (2020). Major histocompatibility complex-associated odour preferences and human mate choice: near and far horizons. *Philos. T. R. Soc. B* 375:20190260. doi: 10.1098/rstb.2019.0260
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X
- Imam, N., and Cleland, T. A. (2020). Rapid online learning and robust recall in a neuromorphic olfactory circuit. *Nat. Mach. Intell.* 2, 181–191. doi: 10.1038/s42256-020-0159-4
- Iravani, B., Arshamian, A., Ravia, A., Mishor, E., Snitz, K., Shushan, S., et al. (2020). Relationship between odor intensity estimates and COVID-19 prevalence prediction in a Swedish population. *Chem. Senses* 22:bjaa034. doi: 10.1093/chemse/bjaa034
- Karlson, P., and Lüscher, M. (1959). “Pheromones”: a new term for a class of biologically active substances. *Nature* 183, 55–56.
- Keller, A., Gerkin, R. C., Guan, Y., Dhurandhar, A., Turu, G., Szalai, B., et al. (2017). Predicting human olfactory perception from chemical features of odor molecules. *Science* 355, 820–826. doi: 10.1126/science.aal2014
- Kennedy, D., and Norman, C. (2005). So much more to know. *Science* 309, 78–102.
- Khan, R. M., Luk, C. H., Flinker, A., Aggarwal, A., Lapid, H., Haddad, R., et al. (2007). Predicting odor pleasantness from odorant structure: pleasantness as a reflection of the physical world. *J. Neurosci.* 27, 10015–10023. doi: 10.1523/JNEUROSCI.1158-07.2007
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cogn. Emot.* 24, 1377–1388. doi: 10.1080/02699930903485076
- Lebanov, L., Tedone, L., Ghiasvand, A., and Paull, B. (2020). Random Forests machine learning applied to gas chromatography–Mass spectrometry derived average mass spectrum data sets for classification and characterisation of essential oils. *Talanta* 208:120471. doi: 10.1016/j.talanta.2019.120471
- Loos, H. M., Doucet, S., Soussignan, R., Hartmann, C., Durand, K., Dittrich, R., et al. (2014). Responsiveness of human neonates to the odor of 5 α -androst-16-en-3-one: a behavioral paradox? *Chem. Senses* 39, 693–703.
- Loos, H. M., Reger, D., and Schaal, B. (2019). The odor of human milk: its chemical variability and detection by newborns. *Physiol. Behav.* 199, 88–99. doi: 10.1016/j.physbeh.2018.11.008
- Lötsch, J., Kringsel, D., and Hummel, T. (2018). Machine learning in human olfactory research. *Chem. Senses* 44, 11–22. doi: 10.1093/chemse/bjy067
- Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., et al. (2018). Differential coding of perception in the world's languages. *Proc. Natl. Acad. Sci. U.S.A.* 115, 11369–11376. doi: 10.1073/pnas.1720419115
- Majid, A., Speed, L., Croijmans, I., and Arshamian, A. (2017). What makes a better smellier? *Perception* 46, 406–430.
- Makin, J. W., and Porter, R. H. (1989). Attractiveness of lactating females' breast odors to neonates. *Child Dev.* 60, 803–810.
- Mayer-Schönberger, V., and Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Houghton Mifflin Harcourt.
- McGann, J. P. (2017). Poor human olfaction is a 19th-century myth. *Science* 356:eaam7263. doi: 10.1126/science.aam7263
- Misaki, M., Kim, Y., Bandettini, P. A., and Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53, 103–118. doi: 10.1016/j.neuroimage.2010.05.051
- Mitro, S., Gordon, A. R., Olsson, M. J., and Lundström, J. N. (2012). The smell of age: perception and discrimination of body odors of different ages. *PLoS One* 7:e0038110. doi: 10.1371/journal.pone.0038110
- Molnar, C. (2018). *A Guide For Making Black Box Models Explainable*. Available online at: <https://christophm.github.io/interpretable-ml-book> (accessed September 16, 2020).
- Mori, T., and Uchihira, N. (2019). Balancing the trade-off between accuracy and interpretability in software defect prediction. *Empir. Softw. Eng.* 24, 779–825. doi: 10.1007/s10664-018-9638-1
- Mujica-Parodi, L. R., Strey, H. H., Frederick, B., Savoy, R., Cox, D., Botanov, Y., et al. (2009). Chemosensory cues to conspecific emotional stress activate amygdala in humans. *PLoS One* 4:e6415. doi: 10.1371/journal.pone.0006415
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.

- Natsch, A., and Emter, R. (2020). The specific biochemistry of human axilla odour formation viewed in an evolutionary context. *Philos. Trans. R. Soc. Lond. Ser. B* 375:20190269. doi: 10.2210/PDB6SLF/PDB
- Olsson, M. J., Lundström, J. N., Kimball, B. A., Gordon, A. R., Karshikoff, B., Hosseini, N., et al. (2014). The scent of disease: human body odor contains an early chemosensory cue of sickness. *Psychol. Sci.* 25, 817–823. doi: 10.1177/0956797613515681
- Parma, V., Gordon, A. R., Cecchetto, C., Cavazzana, A., and Lundström, J. N. (2017). “Processing of human body odors,” in *Springer Handbook of Odor*, ed. A. Buettner, (New York, NY: Springer), 963–986. doi: 10.1007/978-3-319-26932-0
- Parma, V., Ohla, K., Veldhuizen, M. G., Niv, M. Y., Kelly, C. E., Bakke, A. J., et al. (2020). More than smell – COVID-19 is associated with severe impairment of smell, taste, and chemesthesis. *Chem. Senses* 20:bjaa041. doi: 10.1093/chemse/bjaa041
- Pause, B. M. (2017). “Human chemosensory communication,” in *Springer Handbook of Odor*, ed. A. Buettner, (New York, NY: Springer), 987–1010.
- Pause, B. M., Adolph, D., Prehn-Kristensen, A., and Ferstl, R. (2009). Startle response potentiation to chemosensory anxiety signals in socially anxious individuals. *Int. J. Psychophysiol.* 74, 88–92. doi: 10.1016/j.ijpsycho.2009.07.008
- Pause, B. M., Storch, D., and Lübke, K. T. (2020). Chemosensory communication of aggression: women’s fine-tuned neural processing of male aggression signals. *Philos. Trans. R. Soc. B Biol. Sci.* 375:20190270. doi: 10.1098/rstb.2019.0270
- Penn, D. J., Oberzaucher, E., Grammer, K., Fischer, G., Soini, H. A., Wiesler, D., et al. (2007). Individual and gender fingerprints in human body odour. *J. R. Soc. Interface* 4, 331–340. doi: 10.1098/rsif.2006.0182
- Porter, J., Craven, B., Khan, R. M., Chang, S.-J., Kang, I., Judkewitz, B., et al. (2007). Mechanisms of scent-tracking in humans. *Nat. Neurosci.* 10, 27–29. doi: 10.1038/nn1819
- Porter, R. H., Makin, J. W., Davis, L. B., and Christensen, K. M. (1991). An assessment of the salient olfactory environment of formula-fed infants. *Physiol. Behav.* 50, 907–911. doi: 10.1016/0031-9384(91)90413-I
- Prehn-Kristensen, A., Wiesner, C., Bergmann, T. O., Wolff, S., Jansen, O., Mehdorn, H. M., et al. (2009). Induction of empathy by the smell of anxiety. *PLoS One* 4:e5987. doi: 10.1371/journal.pone.0005987
- Preti, G., Dalton, P., and Maute, C. (2019). Analytical identification of stress odors from human breath. *HDIAC J.* 6, 10–15.
- Quintana, P., Nolet, K., Baus, O., and Bouchard, S. (2019). The effect of exposure to fear-related body odorants on anxiety and interpersonal trust toward a virtual character. *Chem. Senses* 44, 683–692. doi: 10.1093/gerona/gly169/5057054
- Regenbogen, C., Axelsson, J., Lasselin, J., Porada, D. K., Sundelin, T., Peter, M. G., et al. (2017). Behavioral and neural correlates to multisensory detection of sick humans. *Proc. Natl. Sci. U.S.A.* 114, 6400–6405. doi: 10.1073/pnas.1617357114
- Reichenbach, S. E., Zini, C. A., Nicolli, K. P., Welke, J. E., Cordero, C., and Tao, Q. (2019). Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition. *J. Chromatogr. A* 1595, 158–167. doi: 10.1016/j.chroma.2019.02.027
- Roberts, S. C., Havlíček, J., and Schaal, B. (2020). Human olfactory communication: current challenges and future prospects. *Philos. Trans. R. Soc. B* 375:20190258. doi: 10.1098/rstb.2019.0258
- Sanchez-Lengeling, B., Wei, J. N., Lee, B. K., Gerkin, R. C., Aspuru-Guzik, A., and Wiltschko, A. B. (2019). *Machine Learning for Scent: Learning Generalizable Perceptual Representations of Small Molecules*. Available online at: <http://arxiv.org/abs/1910.10685> (accessed September 16, 2020).
- Schaal, B., and Porter, R. H. (1991). “Microsmatic humans” revisited: the generation and perception of chemical signals. *Adv. Study Behav* 20, 135–199. doi: 10.1016/S0065-3454(08)60321-6
- Schaal, B., Saxton, T. K., Loos, H., Soussignan, R., and Durand, K. (2020). Olfaction scaffolds the developing human from neonate to adolescent and beyond. *Philos. Trans. R. Soc. B Biol. Sci.* 375:20190261. doi: 10.1098/rstb.2019.0261
- Simons, D. J., Shoda, Y., and Lindsay, D. S. (2017). Constraints on Generality (COG): a proposed Addition to All Empirical Papers. *Perspect. Psychol. Sci.* 12, 1123–1128. doi: 10.1177/1745691617708630
- Singh, P. B., Young, A., Lind, S., Leegaard, M. C., Capuozzo, A., and Parma, V. (2018). Smelling anxiety chemosignals impairs clinical performance of dental students. *Chem. Senses* 43, 411–417. doi: 10.1093/chemse/bjy028
- Smeets, M. A. M., Rosing, E. A. E., Jacobs, D. M., van Velzen, E., Koek, J. H., Blonk, C., et al. (2020). Chemical fingerprints of emotional body odor. *Metabolites* 10:84. doi: 10.3390/metabo10030084
- Snitz, K., Perl, O., Honigstein, D., Secundo, L., Ravia, A., Yablunka, A., et al. (2019). Smellspace: an odor-based social network as a platform for collecting olfactory perceptual data. *Chem. Senses* 44, 267–278. doi: 10.1093/chemse/bjz014
- Snitz, K., Yablunka, A., Weiss, T., Frumin, I., Khan, R. M., and Sobel, N. (2013). Predicting odor perceptual similarity from odor structure. *PLoS Comput. Biol.* 9:e3184. doi: 10.1371/journal.pcbi.1003184
- Sorokowski, P., Karwowski, M., Misiak, M., Marczak, M. K., Dziekan, M., Hummel, T., et al. (2019). Sex differences in human olfaction: a meta-analysis. *Front. Psychol.* 10:242. doi: 10.3389/fpsyg.2019.00242
- Stevenson, R. J. (2010). An initial evaluation of the functions of human olfaction. *Chem. Senses* 35, 3–20. doi: 10.1093/chemse/bjp083
- Stilo, F., Liberto, E., Reichenbach, S. E., Tao, Q., Bicchi, C., and Cordero, C. (2019). Untargeted and targeted fingerprinting of extra virgin olive oil volatiles by comprehensive two-dimensional gas chromatography with mass spectrometry: challenges in long-term studies. *J. Agric. Food Chem.* 67, 5289–5302. doi: 10.1021/acs.jafc.9b01661
- Trivedi, D. K., Sinclair, E., Xu, Y., Sarkar, D., Walton-Doyle, C., Liscio, C., et al. (2019). Discovery of volatile biomarkers of Parkinson’s disease from sebum. *ACS Cent. Sci.* 5, 599–606. doi: 10.1021/acscentsci.8b00879
- Williams, J., Stöner, C., Wicker, J., Krauter, N., Derstroff, B., Bourtsoukidis, E., et al. (2016). Cinema audiences reproducibly vary the chemical composition of air during films, by broadcasting scene specific emissions on breath. *Sci. Rep.* 6:25464. doi: 10.1038/srep25464
- Wyatt, T. D. (2015). The search for human pheromones: the lost decades and the necessity of returning to first principles. *Proc. R. Soc. B Biol. Sci.* 282:20142994. doi: 10.1098/rspb.2014.2994
- Wyatt, T. D. (2020). Reproducible research into human chemical communication by cues and pheromones: learning from psychology’s renaissance. *Philos. Trans. R. Soc. B* 375:20190262. doi: 10.1098/rstb.2019.0262
- Wysocki, C. J., and Gilbert, A. N. (1989). National geographic smell survey: effects of age are heterogeneous. *Ann. N.Y. Acad. Sci.* 561, 12–28. doi: 10.1111/j.1749-6632.1989.tb20966.x
- Yeshurun, Y., and Sobel, N. (2010). An odor is not worth a thousand words: from multidimensional odors to unidimensional odor objects. *Annu. Rev. Psychol.* 61, 219–241. doi: 10.1146/annurev.psych.60.110707.163639
- Zarzo, M. (2011). Hedonic judgments of chemical compounds are correlated with molecular size. *Sensors* 11, 3667–3686. doi: 10.3390/s110403667
- Zhou, W., and Chen, D. (2009). Fear-related chemosignals modulate recognition of fear in ambiguous facial expressions. *Psychol. Sci.* 20, 177–183. doi: 10.1111/j.1467-9280.2009.02263.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 de Groot, Croijmans and Smeets. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.