Methods

A new conceptual and quantitative approach to exploring and defining potential open-access olfactory information

Cristian Gabriel Orlando[1], Malcolm Possell[1], Catherine Price[1], Peter B. Banks[1], Louis Mercorelli[2] and Clare McArthur[1]

[1]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia; [2]The Sydney Informatics Hub, The University of Sydney, Sydney, NSW 2006, Australia

Author for correspondence:
Cristian Gabriel Orlando
Email: cristian.orlando@sydney.edu.au

Received: *29 April 2022*
Accepted: *27 July 2022*

New Phytologist (2022) **236:** 1605–1619
doi: 10.1111/nph.18432

Key words: cue, ecological interactions, information recovery, signal, volatile organic compounds.

Summary

- All organisms emit odour, providing 'open-access' olfactory information for any receiver with the right sensory apparatus. Characterizing open-access information emitted by groups of organisms, such as plant species, provides the means to answer significant questions about ecological interactions and their evolution.
- We present a new conceptual framework defining information reliability and a practical method to characterize and recover information from amongst olfactory noise.
- We quantified odour emissions from two tree species, one focal group and one outgroup, to demonstrate our approach using two new R statistical functions. We explore the consequences of relaxing or tightening criteria defining information and, from thousands of odour combinations, we identify and quantify those few likely to be informative.
- Our method uses core general principles characterizing information while incorporating knowledge of how receivers detect and discriminate odours. We can now map information in consistency–precision reliability space, explore the concept of information, and test information–noise boundaries, and between cues and signals.

Introduction

An organism's fitness depends intimately on how it interacts with the surrounding world. Fitness is almost invariably enhanced if its interactions are nonrandom, based on information garnered from its environment. Plants, herbivores, predators, prey, pollinators, and seed dispersers can benefit from information, using it to make decisions such as where to move (Spiegel & Crofoot, 2016), what flower to visit (Campbell *et al.*, 2010) or what to eat (McArthur *et al.*, 2019), whether to flee from predators (Zuberbühler *et al.*, 1997) or defend from herbivory (Karban *et al.*, 2006).

Among the senses organisms use to acquire information from the environment, olfaction is one of the oldest and most widespread (Eisthen, 2002; Ache & Young, 2005). Odours can travel far from the source and linger in the environment (Doty, 1986), providing information in an odour landscape (Finnerty *et al.*, 2022) long after acoustic and visual information disappear. But of all the compounds that make up an odour, what parts comprise information and how can we define and detect it? To date, the study of the 'ecology of information' (Schmidt *et al.*, 2010) in relation to odour has often focused on defining olfactory information through the lens of specific receivers. Testing the responses of insects (Fraser *et al.*, 2003; Conchou *et al.*, 2017) and herbivorous mammals (Bedoya-Pérez *et al.*, 2014; Schmitt *et al.*, 2018) to odour from plants, for example, has helped define olfactory cues that particular animal species use to find their food. However, every organism emits a wealth of open-access olfactory information, information that is available for any organism to receive, access, and exploit (Vos *et al.*, 2006; Banks *et al.*, 2016; Aljbory & Chen, 2018) provided they possess the appropriate sensory capacity. The limitation of defining odour information from the perspective of a specific receiver is that we necessarily preclude any understanding of the wider set of possible interactions involving the emitter (Fig. 1).

Greater knowledge of open-access olfactory information would allow us to test whether and which of its components are exploited by different receivers within and among ecological communities, to better understand the strategies emitters use to advertise or hide themselves from other organisms, and to obtain insight into the evolution of communication. For example, selective pressure should enhance communication if both the emitter and receiver benefit from their interaction (Schiestl & Johnson, 2013) but lead to an information arms race (deteriorating communication) if the interaction is harmful to either (Zu *et al.*, 2020). Given the complexity of odour profiles emitted from many organisms and how different receivers interpret them (Auffarth, 2013; Tromelin, 2016), characterizing the full suite of information emitted by an organism has been elusive.

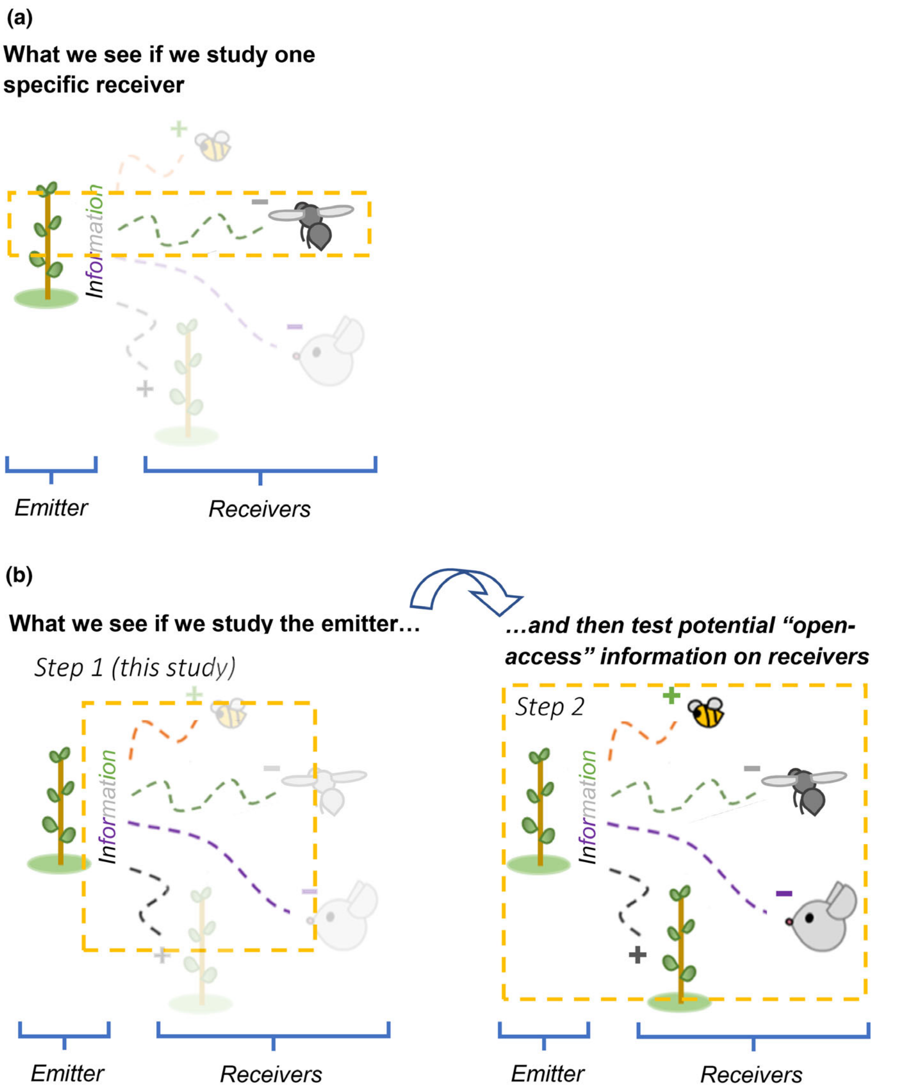

Fig. 1 Conceptual diagram of information flow between an emitter (a plant) and various receivers (such as vertebrate and invertebrate pollinators, herbivores, and other plants). (a) By studying one specific receiver, we likely see only a fraction of both the available information and the possible ecological interactions. (b) By studying the emitter, we can identify all the potential 'open-access' information it produces (step 1 – this study) and then test it on different receivers to detect possible interactions (step 2). Dashed coloured lines indicate the different components of information used by particular receivers, and signs indicate resulting beneficial (+) or harmful (−) interactions for the emitter.

Defining odour information

The odour emitted by organisms comprises a complex mix of volatile organic compounds (VOCs). By definition, information reduces uncertainty about the object of interest (Inglis *et al.*, 2001); for example, the presence or absence of a particular plant species in a vegetation patch. By definition, information is therefore reliable (Koops, 2004) providing a predictable contrast against a background of 'noise' such as VOCs produced by other organisms. Thus, not all of the VOCs produced by an organism will be informative, as many are unreliable. Two main hypotheses have been proposed to explain how VOC information, defining a group of emitters (e.g. a species), is conveyed to receivers. Information is either in the form of unique VOCs or it occurs as specific combinations of nonunique VOCs (Bruce *et al.*, 2005). Unique VOCs involved in interspecific interactions are rare but do exist (Raguso, 2008); for example, the ethyl (2*E*,4*Z*)-2,4-decadienoate kairomone emitted by ripe Bartlett pear (*Pyrus communis*) (Knight & Light, 2001). However, many organisms from different species emit VOCs in common. These shared VOCs can only convey information about a specific group if they are produced in distinct combinations. Such specific combinations have usually been described in terms of ratios of VOC pairs (Webster *et al.*, 2010a).

Two broad approaches have been taken to define olfactory information and detect VOC patterns: biological and statistical. The biological approach defines information by testing the response of a specific receiver organism to a set of VOCs. For example, electroantennography uses the physiological response of specific insects to detect odour compounds 'of interest' to those insects (Hoballah *et al.*, 2005; Riffell *et al.*, 2009), and behavioural assays measure the response of animals such as mammalian herbivores to food odours (Hodgkison *et al.*, 2007; Schmitt *et al.*, 2018). The statistical approach uses a variety of mathematical algorithms to find compounds that are similar among individuals within a group – such as correlation and principal component analysis (Webster *et al.*, 2010b) – or to define a group based on differences with emissions of other groups (such as permutational multivariate ANOVA) (Schmitt *et al.*, 2020).

Box 1 Statistical approaches to defining odour information: the random forest algorithm

The random forest (RF) algorithm is a statistical method often used to analyse odour information and identify compounds specific to particular groups (Ranganathan & Borges, 2010). The RF is a classification method; as such, it is an algorithm that compares groups to find a set of individual volatile organic compounds (VOCs) that distinguish between them. There are two reasons why this approach is limited:

(**1**) Relevant information is missed: the RF method works by finding the smallest number of VOCs that can differentiate and classify different groups. Therefore, some relevant VOCs in a group's odour profile can be excluded either because classification is reached with a smaller set of VOCs or because some VOCs are shared between groups. The RF method does not deal with combinations of VOCs nor provide us with the specific ratio between compounds. Yet, biological evidence shows that invertebrate receivers, for example, often use combinations of VOCs defined as specific ratios (i.e. relative proportions) as information (Bruce & Pickett, 2011). Furthermore, VOCs shared among groups could still be informative; for example, if they occur in distinct combinations including distinct relative proportions.

(**2**) Results depend entirely on an outgroup: when looking for differences between groups, the specific group information can change depending on the set of groups used. Thus, the addition or exclusion of groups with similar characteristics can lead to changes in the list of VOCs that are useful to classify groups.

Both approaches help define some informative odour compounds. The biological approach allows us to identify information used by specific receivers. From this approach, we know that, within certain plant species, VOC pairs in specific ratios can be relatively consistent among individuals and that insects respond to the average of such ratios (Webster *et al.*, 2010b). The extent to which these VOC pairs, and their ratios, can vary among individual plants yet still be informative to different receivers is unknown. The statistical approach helps us to identify VOCs in common within a group and/or distinct from other groups from across the full range of odour emissions of organisms. However, different receivers can use the same or different subsets of the entire information 'package' emitted by an organism (Beyaert & Hilker, 2014; Kessler & Kalske, 2018), depending on their sensory apparatus; and the subsets they use are not necessarily the same as those detected using statistical methods (see Box 1 for further explanation).

Here, we propose a new, flexible approach to defining potential open-access information that complements the biological and statistical approaches. Using our approach, biologists can quantitatively explore the potential range of information within odour profiles, part or all of which could be used by a variety of receivers to identify an emitter. It also provides the practical basis for exploring boundaries between information and noise, and between cues (information that is provided unintentionally by the emitter) and signals (information produced for an intentional communication) (Steiger *et al.*, 2010; Ninkovic *et al.*, 2021) in ecological interactions. In the following sections, we introduce and explain the conceptual framework underpinning the approach and demonstrate how it can be used.

Conceptual framework, approach and aims

Our conceptual framework (Fig. 2) is that odour information characterizing any group (such as an organ, species, population, or phenological phase) is defined by reliability, and that reliability increases along two within-group axes that we term consistency and precision. Consistency is the proportion of individuals emitting the same VOC combination, such as the same pair of VOCs. Precision is the degree to which the relative amounts of VOCs in a particular combination (e.g. a VOC pair) vary among individuals that have that combination. Thus, the more individuals within a group emitting the same VOC combination, the more consistent and hence more reliable is the consistency trait within that group. Similarly, the less variation among individuals in the relative amounts (e.g. ratio) of VOCs in a particular combination, the more precise and, hence, the more reliable is this precision trait. Volatile organic compound combinations that have both low consistency and low precision are uninformative noise. We hypothesize that information in the form of cues – passive traits such as odours released incidentally (Ninkovic *et al.*, 2021) – may be relatively consistent or relatively precise but rarely both, as they are, by definition, not intentionally produced nor under selection pressure to provide information, whereas signals – released with the function of actively communicating with a receiver (Ninkovic *et al.*, 2021) – will be both highly consistent

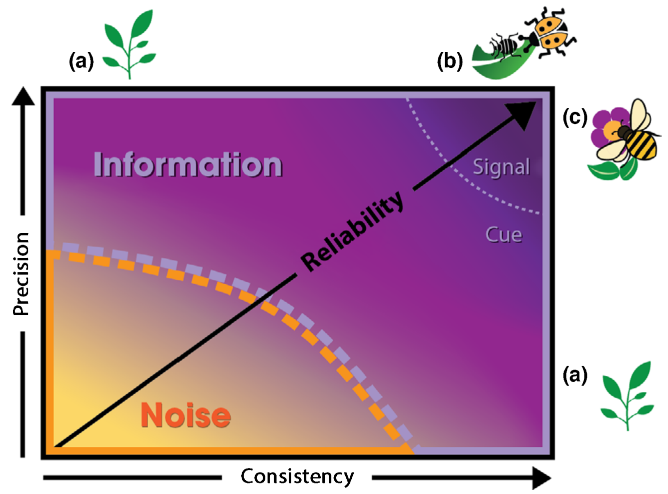

Fig. 2 Conceptual framework showing how consistency and precision, two within-group criteria, relate to noise vs information; and within information, cue vs signal. Reliability of odour emissions increases with increasing consistency (occurring in more members of a group) and/or precision (tight relative amounts of the volatile organic compounds (VOCs) among members of a group). Odour emissions with both low consistency and low precision are completely unreliable, uninformative, and hence noise. Cues, such as (a) odours emitted from undamaged palatable seedlings, are predicted to have VOC combinations that are either highly consistent or highly precise but rarely both. Signals, such as plant odour emissions aimed to (b) attract natural enemies of damaging insect herbivores or (c) attract pollinators should have both highly consistent and highly precise VOC combinations.

and highly precise. Beneficial cues – information produced unintentionally that is used by a receiver subsequently benefiting the emitter – can be evolutionarily selected and enhanced, thus becoming a signal (Steiger *et al.*, 2010). Unfavourable cues will likely enter an arms race, maintaining intermediate levels of reliability (Zu *et al.*, 2020). In some cases, unwanted receivers (e.g. a forager) may eavesdrop on signals communicating to other receivers, but that does not change the fact that the signal is 'intentionally' produced.

This conceptual framework underpins our approach for quantitatively identifying potentially informative combinations of nonunique VOCs within the emissions of any focal group. The approach uses a simple sequential procedure based on the two within-group criteria, consistency and precision, followed by a between-group criterion, termed accuracy. Volatile organic compound combinations are more accurate at defining a group when they differ from those emitted by other groups. These three criteria together constitute undisputed logical attributes of reliable olfactory information. By considering these three criteria consecutively, our heuristic method allows us to characterize and quantify the potential open-access olfactory information of emitters within a focal group – the group's signature or fingerprint – and it does so independent of a receiver.

Our method first defines all potentially informative VOCs and their relative amounts within the focal group using the consistency and precision criteria. It focuses on the relative amounts of VOCs to incorporate the biological evidence (Visser & Avé, 1978; Birkett *et al.*, 2004; Beyaert *et al.*, 2010; Bruce & Pickett, 2011) that receivers frequently have to rely on relative amounts (usually quantified as ratios) of nonunique VOCs for information, rather than relying on unique VOCs. Accuracy then tests for differences between the focal group and other groups (outgroups) to help confirm whether the potentially informative within-group set of VOCs differs from those in other groups in the landscape. This accuracy step is analogous to that used in typical biological and statistical approaches, with the same limitation that the result necessarily depends on which, and how many, outgroups are tested. Using this conceptual framework and heuristic method, we then had two aims.

Our first aim was to explore the consequences of relaxing or tightening the first two criteria (consistency and precision) for defining within-group information. In other words, how many and which VOC combinations enter or leave the information set as a function of the specific values of consistency and precision? This aim is important because – even though conceptually information comprises reliable components of odours, differentiating them from within-species noise (Fig. 2) – there is no single objective level at which this boundary can be defined.

Our second aim was to demonstrate how to quantitatively identify potentially informative combinations of nonunique odour compounds and their associated relative amounts, using all three criteria.

To achieve both aims, we generated a data set of odour emissions from seedlings of two tree species, one focal group and one outgroup, and developed two new functions in RStudio (RStudio, 2015) to automate the process for future users.

Finally, to better demonstrate the conceptual framework and the heuristic method, we compared our results with those obtained by the RF algorithm.

Study system

Our focal group comprised seedlings of the tree species *Eucalyptus punctata*, and our outgroup comprised seedlings of the tree species *Corymbia gummifera* (formerly *Eucalyptus gummifera*). *Eucalyptus punctata* grows in many woodlands and forests of eastern Australia (Boland *et al.*, 2006), and the leaves are an important food source for a variety of vertebrates and invertebrates (Cork *et al.*, 1983; Steinbauer, 1996; Saunders & Burgin, 2001). Leaves of *E. punctata* contain many terpenes (Southwell, 1973; Dellacassa *et al.*, 1990) that give them a distinct odour. The dominant native mammalian herbivore in many of these ecosystems, the swamp wallaby *Wallabia bicolor*, uses leaf odour as a cue to find *E. punctata* seedlings efficiently and nonrandomly (Finnerty *et al.*, 2017). We have yet to define the specific informative odour components from within the whole odour profile of *E. punctata* that wallabies, or indeed any other receiver, could rely on as a cue for this palatable species. We chose *C. gummifera* as an appropriate and relevant outgroup because both *E. punctata* and *C. gummifera* are 'eucalypts', hence likely sharing similar odour compounds, and they are also often sympatric. From a receiver perspective, distinguishing between different eucalypt species is important as they will have preferences associated with leaf chemistry and structure (Wiggins *et al.*, 2006; Gallahar *et al.*, 2021).

Importantly, we analysed the odour profiles of undamaged (nonflowering) seedlings. We wanted to focus on odours we predicted were cues rather than signals, to provide a greater challenge for testing our method. Signals, by definition, should have high reliability (Fig. 2), and so should be easy to detect; and many plants damaged by herbivores emit signals. For example, some plants emit induced VOCs, such as salicylic acid, jasmonic acid, or specific sets of VOCs, a 'cry for help' that attracts natural enemies of insect herbivores, or that conspecific neighbouring plants detect, leading to their own induced chemical defences (Aljbory & Chen, 2018; Ninkovic *et al.*, 2021). Undamaged leaves, on the other hand, may incidentally emit constitutive VOCs with informative components characteristic of the group, but they are less likely to emit a signal because of the risk of attracting herbivores. Such cues may be less reliable than a signal (Fig. 2), either because selective pressure for similarity within the group is relaxed or because individuals within the group are under selective pressure to avoid herbivory by being different.

Materials and Methods

Plant material and sampling conditions

Plant VOC emissions are a function of plant genotype, the circadian cycle, and abiotic factors such as temperature (Zini *et al.*, 2002; Hoballah *et al.*, 2005; Beyaert & Hilker, 2014) and even biotic factors such as herbivory (Low *et al.*, 2014; Aartsma *et al.*, 2017). However, if these emissions are to convey information about individuals within a group, such as a species, some of these emissions must be *consistent* and *precise* irrespective of this variation. To detect information emitted by our focal group of *E. punctata* while including phenotypic variability, we sampled 38 seedlings (*c.* 10 months old) commercially sourced at different times (possibly of different genetic stock) in three conditions:

(**1**) Batch 1: headspace samples of six *E. punctata* seedlings were taken in a glasshouse during daytime (between 08:00 and 17:00 h local time (LT)), exposed to summer natural light and temperature (23.5–38.5°C) variation on a sunny day.

(**2**) Batch 2: headspace samples of 22 *E. punctata* seedlings were taken in a glasshouse during daytime (between 09:00 and 17:00 h LT), exposed to winter natural light and temperature (15–22°C) variation on a cloudy day.

(**3**) Batch 3: headspace samples of 10 *E. punctata* seedlings were taken in a laboratory under controlled light (300 μmol m^{-2} s^{-1} photosynthetically active radiation; 20 W LED light; Arlec Pty. Ltd, Blackburn North, Vic., Australia) and temperature (21–25°C) conditions (between 09:00 and 17:00 h LT).

To then test for accuracy (i.e. species-specific information), we compared the information emitted by *E. punctata* seedlings against *C. gummifera* seedlings. Headspace samples of 11 *C. gummifera* seedlings were taken under the same conditions and at the same time as the corresponding *E. punctata* from batch 3.

VOC sampling

The odour profile of each individual seedling in batch 1 and batch 3 was measured using a branch enclosure (for more information, see Supporting Information Methods S1) and in batch 2 using heat-resistant polyester bags – which increases the concentration of VOCs within the bag in comparison with the other method. These techniques allowed us to measure emissions from live, undamaged plants and to detect VOCs of low concentration or short lifespan – such as caryophyllene, which could degrade shortly after release (Calogirou *et al.*, 1997). Volatile organic compounds were collected from the enclosures using a sorbent tube (Tenax TA; Markes International Ltd, Bridgend, UK) connected to an air pump (AirChek 2000; SKC Inc., Eighty-Four, PA, USA; or PAS-500; Spectrex Corp., Redwood City, CA, USA) flowing air at 200 ml min^{-1} for 30 min. A 10–30 min gap between setting up each sample and sampling was used to allow the air inside the branch enclosure (or polyester bag) to reach an equilibrium state. Background samples (VOCs in the air in the empty branch enclosure) were also taken for each batch of *E. punctata* and *C. gummifera*. Control (blank) samples were taken at the beginning (08:00 h) and the end (16:30 h) of the sampling session for batch 1, and with each seedling sampled in batch 2 and batch 3. Post-sampling, tubes were maintained at 4°C until analysis by GC–MS (within 2 wk from sampling).

GC–MS analysis

Sorbent tubes were analysed by thermal desorption GC–MS using the same equipment and protocols described in Lawson *et al.* (2020) (for more information, see Methods S1). Postprocessing of the GC–MS data was performed using unsupervised data mining with the package MSEASY (v.5.5.3; Nicolè *et al.*, 2012) in R (v.3.5.2; R Core Team, 2019). This program is insensitive to shifts in retention times and detects putative compounds within complex metabolic mixtures through the clustering of mass spectra. Retention times, silhouette width (Rousseeuw, 1987), and Dunn's index (Dunn, 1974) were used for quality control of putative compounds. MSEASY also produced peak areas for the putative compounds based upon their total ion count (TIC). Final ion counts of the putative compounds emitted by our plants were obtained by subtraction of the background TIC recorded for each compound from the plant samples. The most likely identity of each putative compound was determined using a combination of the library's calculated match factor (NIST, 2008), visual comparison of mass spectra with the NIST08 database (NIST, 2008) and that of Kovat's nonisothermal retention indexes, as described in van den Dool & Kratz (1963), calculated using a series of C_8 to C_{20} homologous alkanes. The final ion count of those putative compounds identified as the same compound was added up to obtain only one value per compound.

Defining odour information using the heuristic approach

For our first aim – exploring the consequences of relaxing or tightening the two within-group criteria (consistency and precision) for defining within-group information – we used the emission data from our focal *E. punctata* group. With the new R function ('Odour Information Exploration Algorithm', see Notes S1), we varied the specific values of these criteria to generate a matrix of the number of potentially informative VOC pairs as a function of these values.

Step 1: For the criterion of consistency (VOC pairs present in the majority of the seedlings), we tested a range of values from 0.5 (present in half the seedlings) to 1.0 (present in all seedlings), with values closer to 1.0 representing higher consistency. We used pairs, as often used in the biological approach, because it is a neat and computationally simple way to incorporate the interdependence of multiple VOCs in providing information.

Step 2: For the criterion of precision (variability among seedlings in the relative amount of two VOCs within pairs, from step 1), we tested a range from 0 to 0.5 (out of a total of one), with values closer to zero representing higher precision. For this criterion, we only used replicates that emitted both compounds within a pair. To calculate relative amounts of two VOCs within a pair, we did not quantify them as ratios (dividing relative abundances of two VOCs) as typically used in the biological approach. Ratios are problematic because values above and below 1 cannot be compared (variation is highly distorted with 0 to 1 vs 1 to infinity). Instead, we calculated paired proportions after a fourth-root transformation, by dividing the transformed TIC from one of the VOCs in the pair by the sum of the transformed TIC for both VOCs in the pair. We transformed the data for two reasons: first, animals can be sensitive to compounds in very small concentrations

(Riffell *et al.*, 2009), hence the relationship between many sensory systems and the concentrations of the stimuli approximates to a logarithmic (similar to the fourth root, but the latter allows zeros) dependence (Hopfield, 1999; Hiratani & Latham, 2020); and second, it enabled a better representation of variation among individuals in a group, even if the difference in the relative abundance of two VOCs in a pair was large. Though we use VOCs pairs for analysis of precision in this step, interpretation of results can lead to the identification of higher-order consistent combinations. Thus, for example, if three compounds are precise in all their possible combinations, the information is likely produced as a triplet rather than a pair.

For our second aim – to demonstrate how to quantitatively identify potentially informative combinations of nonunique odour compounds and their associated relative proportions using all three criteria – we chose particular values for consistency and precision based on the output from earlier herein (and see the Results section, Fig. 3). We used these particular values to run the 'Odour Information Definition Algorithm' (Notes S1) and then applied criterion 3 (accuracy). Recognizing there is no single correct level of consistency and precision that objectively distinguishes information from noise, we selected the value of 0.67 for consistency and the value of 0.21 for precision. The choice itself was semi-arbitrary but adequate for our demonstration. It was based on (1) consistency of 67% (two-thirds of seedlings) seemed a reasonable minimum level of reliability to be informative for the group, (2) this consistency combined with 0.21 precision gave four VOCs in four VOC pairs, in line with (3) the biological electroantennogram evidence that insects respond to appropriate blends of a small number of VOCs from a plant species (Fraser *et al.*, 2003; Hoballah *et al.*, 2005; Webster *et al.*, 2008; Riffell *et al.*, 2009).

Defining odour information using the statistical approach

We used the RF algorithm (Breiman, 2001), which creates multiple decision trees that classify the samples into different pre-established groups. The best tree is then selected based on parsimony, indicating the minimum set of relevant attributes that allows samples classification into the different groups. We used the VSURF (Genuer *et al.*, 2015) package to run this analysis in RSTUDIO (RStudio, 2015; Notes S2). For our data, we established two groups: one for seedlings of *E. punctata* and one for seedlings of *C. gummifera*. Analysis of the set of attributes (i.e. VOCs) that are relevant for classification identifies putative odour information emitted by *E. punctata* and *C. gummifera* based only on individual VOCs rather than based on any relative combination of VOCs (such as pairs).

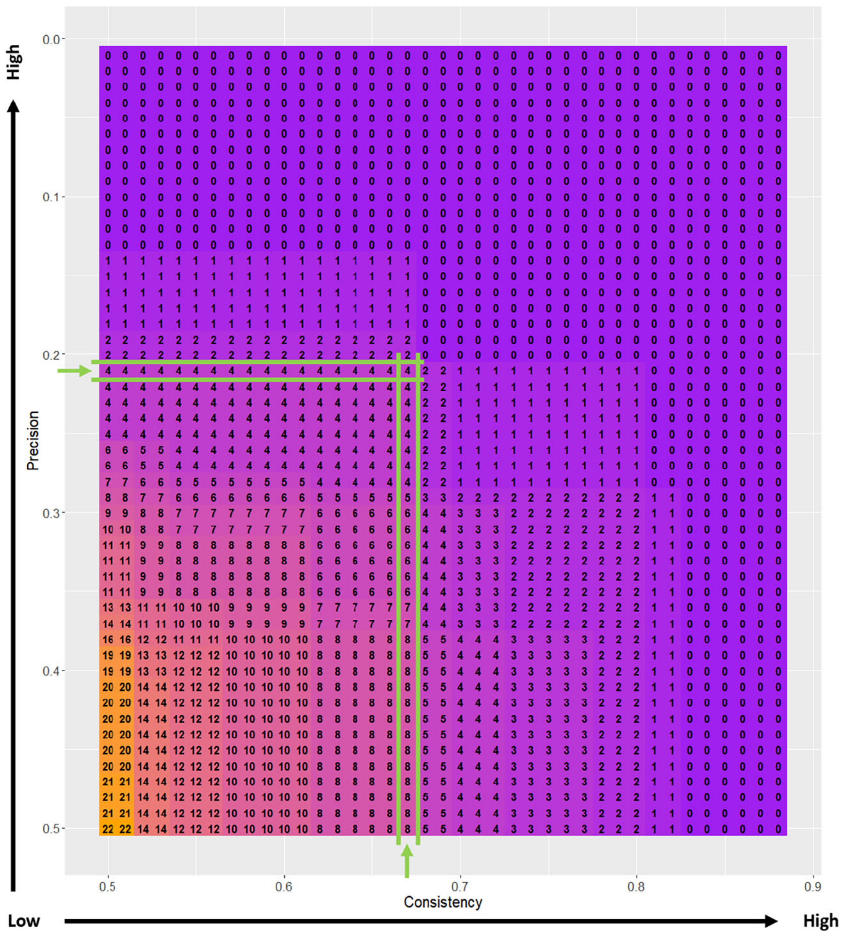

Fig. 3 Matrix of results from the focal group of 38 *Eucalyptus punctata* seedlings using the heuristic method to identify number of potentially informative volatile organic compound (VOC) pairs by varying values for the two within-group criteria, consistency and precision. The numbers in the matrix indicate the number of informative VOC pairs for each consistency and precision combination. The colour of the matrix changes from purple, for a low number of informative VOC pairs, to orange, for a high number of informative VOC pairs. Green arrows indicate the specific values chosen for further investigation in aim 2, in which we applied all three criteria (consistency, precision, and accuracy) to the focal group, *E. punctata*, by including an outgroup, *Corymbia gummifera*.

Results

Odour profiles of *E. punctata* and *C. gummifera* seedlings

After background odour subtraction, we identified a total of 418 likely plant VOCs from the 38 *E. punctata* seedlings (Dataset S1): 45 plant VOCs for batch 1, 366 for batch 2, and 85 for batch 3 (see data in Supporting Information). The difference in number among batches is likely related to differences in sampling techniques and the number of seedlings sampled for each batch. For *C. gummifera*, we identified a total of 72 plant VOCs (Dataset S2).

Defining information using the heuristic approach: consequences of varying within-group criteria

Results from the heuristic method showed that the number of potentially informative pairs in *E. punctata* seedlings varied depending on the level of consistency and precision we specified (Fig. 3) and decreased with both increasing consistency and increasing precision. However, even with highly relaxed criteria (0.5 consistency and 0.5 variation in precision), we were able to reduce a data set of 87 153 possible VOC pair combinations, to identify just 22 of them as potentially informative VOC pairs (Fig. 3). Thus, the first two criteria of our heuristic method defined most VOC emissions (> 99.9% of the VOC pairs) from this plant species as 'noise'. When we tightened the criteria, with a consistency closer to 1 and precision closer to 0, few or no reliable VOC pairs were identified as informative. For example, only one VOC pair was defined as potentially informative when consistency was 0.80 and precision was 0.21 (Fig. 3).

Different combinations of consistency and precision values could define the same number of potentially informative VOC pairs, but, importantly, the identity of these compound pairs was not always the same (Fig. 4). For example, four pairs were identified with consistency of 0.67 and precision of 0.21, and with consistency of 0.70 and precision of 0.40 (Fig. 3). But whereas 1,8-cineole/α-pinene was common to both sets, the three other VOC pairs differed: 1,8-cineole/β-pinene, 1,8-cineole/α-terpineol, and α-pinene/β-pinene in the former set; 1,8-cineole/*p*-cymene, *p*-cymene/α-pinene, and isoprene/*p*-cymene in the latter set (Fig. 4).

Different combinations of consistency and precision values also resulted in variations in the number of individual seedlings that had all, some, or none of the informative VOC pairs identified (Fig. 5). For example, with consistency of 0.67 and precision of 0.21, 50% of the 38 seedlings had all four potentially informative pairs, 16% had three, 13% had two, 5% had one, and 16% of seedlings had none (one from batch 1, three for batch 2, and two for batch 3). By contrast, with a consistency of 0.70 and

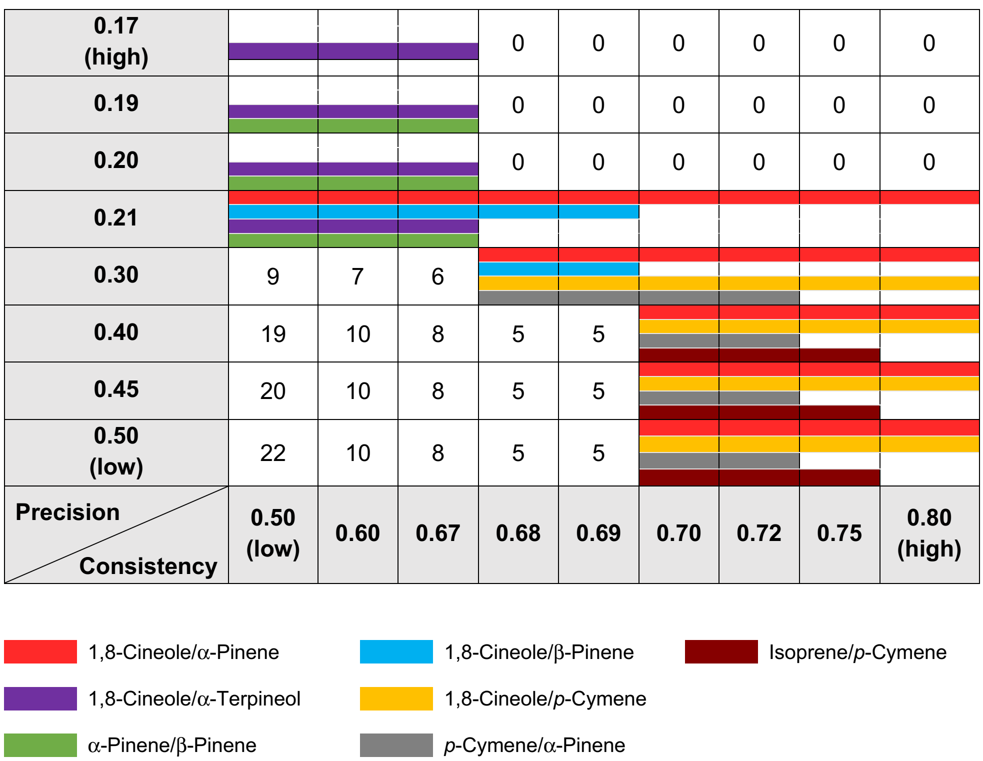

Fig. 4 Results from the focal group of 38 *Eucalyptus punctata* seedlings using the heuristic method to identify pairs of volatile organic compounds (VOCs) in consistency–precision reliability space. Here, values for the two within-group criteria vary: consistency (low 0.5 to high 0.8) and precision (low 0.5 to high 0.17). Specific VOC pairs, presented in a different colour, are shown when one to four pairs were detected. Where a cell contains the number 0, no pairs were identified. Where more than four pairs were identified, just the number of pairs is shown (as in Fig. 2). The method and R function 'Odour Information Definition Algorithm' also provide output on the relative proportion of the two component VOCs in a given pair, but this detailed information is not presented here.

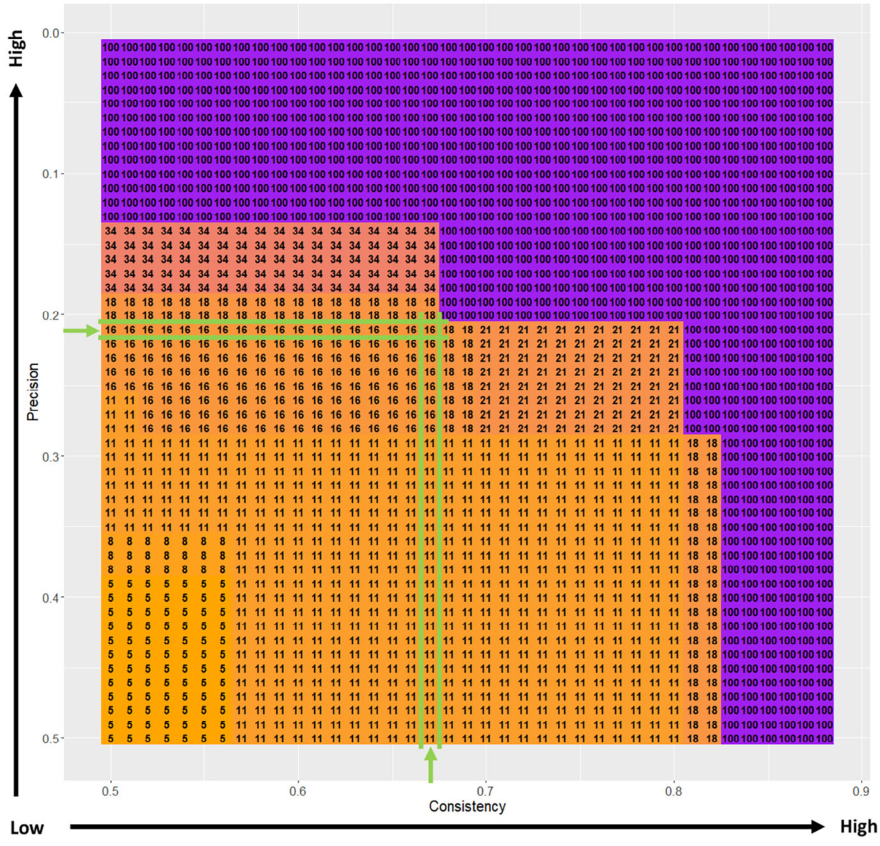

Fig. 5 Matrix generated from the heuristic method showing percentages of the focal group of 38 *Eucalyptus punctata* seedlings that have no informative volatile organic compound (VOC) pairs, for each combination of the two within-group criteria, consistency and precision. The colour of the matrix changes from orange, for a low percentage of replicates without informative VOC pairs, to purple, for a high percentage of replicates without informative VOC pairs. Green arrows indicate the specific values chosen for further investigation in aim 2, in which we applied all three criteria (consistency, precision, and accuracy) to the focal group, *E. punctata*, by including an outgroup, *Corymbia gummifera*.

precision of 0.40, 63% of the 38 seedlings had all four potentially informative pairs, but 11% of seedlings had none.

Defining information using the heuristic approach with all three criteria

Using the same specified values of consistency (0.67) and precision (0.21) for both *E. punctata* and *C. gummifera*, and applying our two within-group criteria, we found the following.

For the *E. punctata* focal group, there were only six consistent VOCs (out of 418) (Fig. 6a): 1,8-cineole, isoprene, *p*-cymene, α-pinene, α-terpineol, and β-pinene. These individual VOCs form 15 possible pairs, from which only four pairs were consistent (criterion 1) and precise (criterion 2) (Fig. 6b): 1,8-cineole/α-pinene, 1,8-cineole/α-terpineol, 1,8-cineole/β-pinene, and α-pinene/β-pinene. As α-pinene and β-pinene formed a precise pair, and both α-pinene and β-pinene each formed a precise pair with 1,8-cineole, they may be part of a higher-order consistent combination (i.e. a triplet).

For the *C. gummifera* outgroup, there were eight consistent VOCs (out of 72): 1-hexanol, 2-ethyl-; acetophenone; butanoic acid, 3-methyl-, 3-methylbutyl ester; isoprene; nonanal; *p*-cymene; α-pinene; β-pinene. Of these, 17 pairs (out of 28 possible pairs) from all eight VOCs were identified as consistent (criterion 1) and precise (criterion 2) (Table 1).

Once we then applied criterion 3 (accuracy) to the datasets, only one pair (α-pinene/β-pinene) in *E. punctata* was also present in *C. gummifera* (Table 1). However, the relative abundance of the two VOCs in this α-pinene/β-pinene pair differed significantly (independent samples *t*-test: $t = 7.19$, $df = 31$, $P < 0.001$) between *E. punctata* (average proportion of 0.62 α-pinene/0.38 β-pinene ± 0.06 (mean ± SD)) and *C. gummifera* (0.52 α-pinene/0.48 β-pinene ± 0.02 (mean ± SD)).

In summary, for the *E. punctata* focal group, three VOC pairs were defined as potentially informative (consistent, precise, and accurate): 1,8-cineole/α-pinene, 1,8-cineole/α-terpineol, and 1,8-cineole/β-pinene. The fourth pair identified by criteria 1 and 2 (α-pinene/β-pinene) was also highly likely to be informative (consistent, precise, and likely accurate), either on its own or as part of a triplet.

Defining information using the RF statistical method

The first step of the RF analysis identified 70 relevant VOCs based on variable importance ranking. These were then reduced to six VOCs by the variable selection for the interpretation step. Finally, the RF analysis indicated that three VOCs were enough to make a perfect prediction of each sample as either *E. punctata* or *C. gummifera*. These VOCs were 1,8-cineole; butanoic acid, 3-methyl-, 3-methylbutyl ester; and β-pinene. Results (Fig. 7) showed that 1,8-cineole was mostly absent from *C. gummifera* (present only in five out of 11 seedlings and with a low ion count); butanoic acid, 3-methyl-, 3-methylbutyl ester was absent in batch 1 and batch 2 of *E. punctata*; and β-pinene was prevalent

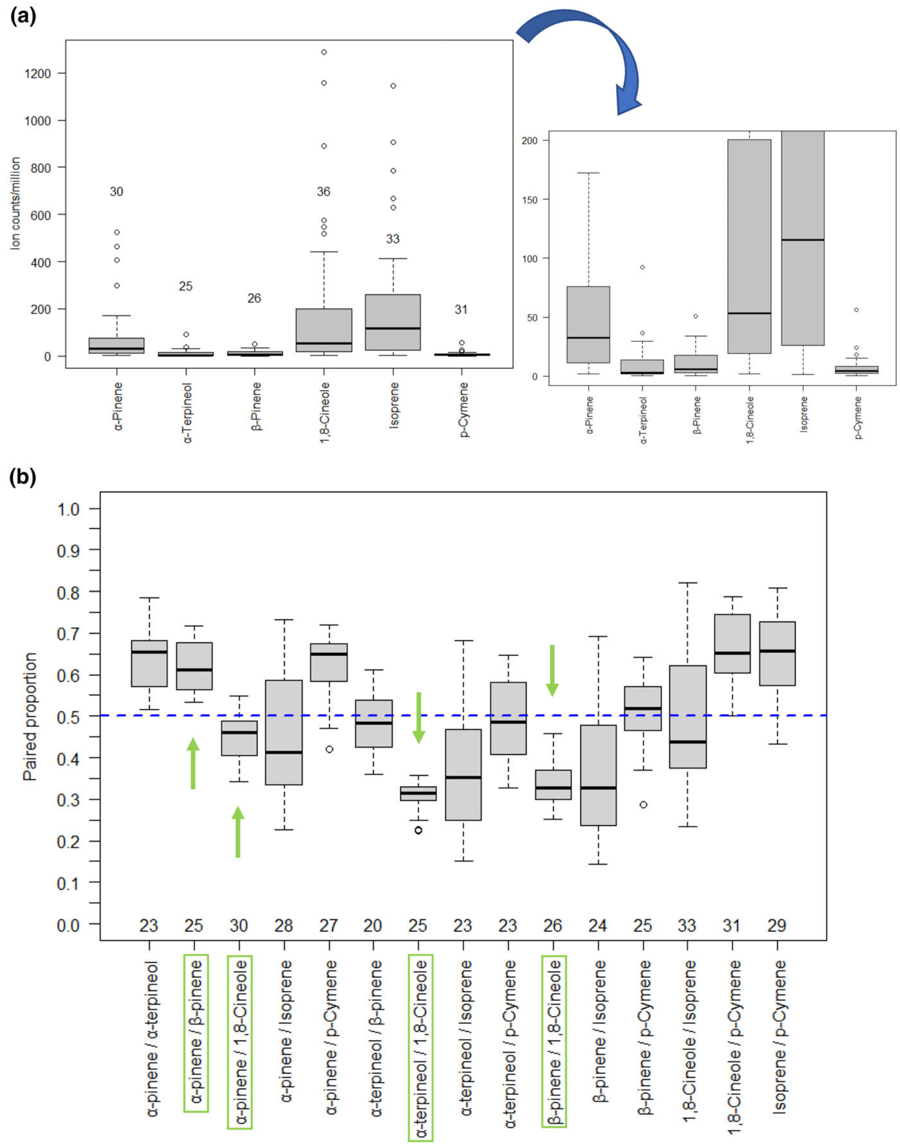

Fig. 6 Results for the focal group of 38 *Eucalyptus punctata* seedlings, using the heuristic method and applying the two within-group criteria. (a) Criterion 1, consistency (0.67 of all seedlings): abundance (as total ion count per million) of the six most frequent volatile organic compounds (VOCs) (when present). Left and right plots show the same information but at a different scale. (b) Criterion 2, precision (variation 0.21): proportion of the first listed compound in each VOC pair (calculated after fourth-root transformation, see the Materials and Methods section) for all 15 possible combinations of the six most frequent VOCs identified in (a) criterion 1. Blue dashed line indicates the space where the paired proportion is 0.5 (i.e. both VOCs in the pair have the same relative abundance). Green arrows indicate the final four precise VOC pairs identified using these specific values for consistency (0.67) and precision (0.21). Boxplots indicate the median, the first and third quartiles, and the maximum and minimum values. Empty circles indicate outliers. Numbers inside the figures indicate the number of seedlings (out of 38) in which (a) each VOC or (b) a pair of VOCs was detected.

in both species but showed a larger ion count for *C. gummifera*. The three compounds selected in the interpretation step but not included in the prediction step were limonene (only present in *C. gummifera*) and *p*-cymene and α-terpineol (present in both plant species but a larger ion count for *E. punctata*).

Discussion

By examining the VOCs emitted by plants and applying a new conceptual framework and quantitative heuristic method based on three criteria for reliability – VOC combinations and relative amounts that are consistent, precise, and accurate – we have characterized those parts of the emission that we argue can define open-access odour information, available for those receivers with the appropriate sensory capacity. This method for recovering quantitative information, without the constraint of using a receiver to do so, can be used to define information emitted by individuals (whole organisms or parts of them, such as leaves, fruits, flowers) of any group, whether groups are considered at the species, population, or subpopulation level, classified by phenological stage (e.g. young and old leaves, or ripe and unripe fruits) or phenotype (e.g. from low- or high-fertile soils, sunny or shady habitat, during night or day). In providing the first logical step of defining the potential information produced by an emitter, ecologists can move on to test whether and what parts different receivers use.

Our demonstration that VOC pairs fall in and out of the potentially informative set as a function of the specific values set for consistency and precision criteria (aim 1) highlights the elusive nature of defining information as distinct from noise. But the flexibility of our method means we can explore the boundaries between information and noise in any given system as we relax or tighten these criteria. This exploration will involve not just defining open-access information vs noise (Fig. 2), as we have demonstrated here, but also any differences between cues vs signals. In doing so, we can start to understand the characteristics of

Table 1 Criterion 3: comparison between the potentially informative volatile organic compounds (VOCs) pairs identified for *Eucalyptus punctata* and *Corymbia gummifera* (indicated by an 'X') when consistency is 0.67 (criterion 1) and precision is 0.21 (criterion 2) for both species.

Potentially informative VOC pairs	*E. punctata*	*C. gummifera*
α-Pinene/1,8-cineole	X	
α-Terpineol/1,8-cineole	X	
β-Pinene/1,8-cineole	X	
α-Pinene/β-pinene	0.62/0.38 (±0.06)	0.52/0.48 (±0.02)
1-Hexanol, 2-ethyl-/butanoic acid, 3-methyl-, 3-methylbutyl ester		X
1-Hexanol, 2-ethyl-/nonanal		X
1-Hexanol, 2-ethyl-/*p*-cymene		X
1-Hexanol, 2-ethyl-/α-pinene		X
Acetophenone/butanoic acid, 3-methyl-, 3-methylbutyl ester		X
Acetophenone/nonanal		X
Acetophenone/*p*-cymene		X
Acetophenone/α-pinene		X
Acetophenone/β-pinene		X
Butanoic acid, 3-methyl-, 3-methylbutyl ester/isoprene		X
Butanoic acid, 3-methyl-, 3-methylbutyl ester/nonanal		X
Butanoic acid, 3-methyl- 3-methylbutyl ester/*p*-cymene		X
Butanoic acid, 3-methyl-, 3-methylbutyl ester/α-pinene		X
Nonanal/*p*-cymene		X
Nonanal/α-pinene		X
p-Cymene/α-pinene		X

For those VOC pairs present in both plant species, the average proportion (±SD) is indicated.

VOC sets (i.e. consistency and precision combinations) that make them informative to different receivers, such as invertebrates, vertebrates, or plants. Depending on their ecology, these receivers may use informative components that occur in different locations in the consistency–precision reliability space. For example, because generalist herbivores are more willing than specialists to search for and feed from a range of plants emitting a variety of information (Vos *et al.*, 2006), they may also be more capable than specialists of recognizing informative VOC pairs that have low precision; that is, large variation in relative proportions ('ratios') of informative compounds among individuals.

In providing an objective, flexible solution to the information recovery problem, we can now test evolutionary and ecological hypotheses for existing questions and ask new questions about the role of olfactory information in ecological interactions. For example, can emitters be both conspicuous (e.g. to beneficial pollinators) and camouflaged (e.g. to harmful foragers) in an odour landscape, if these receivers detect different channels (components) of the overall open-access information?

The fact that the heuristic method defined information for our focal group of *E. punctata* seedlings differently from that of the RF algorithm is not surprising. The two methods take very different approaches. The heuristic method incorporates both the fundamental principles of information and the biological knowledge that information is not simply about individual compounds but about their relative amounts. We argue that this provides the heuristic method with an advantage over the RF algorithm, and other purely statistical methods, when the relative amounts of VOCs matter more than the presence of individual VOCs – for example, when the VOCs identified are commonly emitted by other species in the same habitat. Nevertheless, the RF approach is likely better placed for detecting VOCs that are unique to one group compared with an outgroup.

Distilling information efficiently from amongst the noise

Most of the volatile compounds identified for both *E. punctata* and *C. gummifera* are commonly emitted by plants and are the product of cell membrane damage in growing plants or after mechanical injury (such as hexanal, octanal, and decanal; Hu *et al.*, 2008), or produced as plant secondary metabolites (such as 1,8-cineole and benzoic acid; Dearing *et al.*, 2005; Elaissi *et al.*, 2011), or from the breakdown of other compounds (methyl vinyl ketone is a reaction product of isoprene; Jardine *et al.*, 2012). Therefore, it is highly unlikely that any individual compound can be informative and useful in distinguishing our focal group, *E. punctata*, from others.

Even though the boundary between noise and information remains necessarily blurred, our method provides a powerful and efficient way to hone in on likely informative VOC combinations. By applying the two within-group criteria to the odour profile of *E. punctata* seedlings, we were able to cull a dataset of over 80 000 possible VOC pairs by over 99%; quantifying just 0.025% (22 pairs) as potentially informative even with highly relaxed criteria (0.5 consistency and 0.5 variation in precision) (Fig. 3). This efficiency in data reduction is valuable because it overcomes two practical hurdles in the biological method. First, the need to make an initial guestimate as to what VOC combinations could possibly be informative. Second, having to test and confirm the many possible VOC combinations with animals. Even using electroantennograms with specific insects (Hoballah *et al.*, 2005; Riffell *et al.*, 2009), finding the informative VOC combinations can be time consuming for unfamiliar emitter groups (e.g. new plant species), and is certainly a sophisticated and expensive process.

Our approach will be particularly useful for understanding how vertebrates use information – because the current method of using antennograms for insects is clearly impossible. Instead, future studies can now narrow the putative set of informative VOCs quickly and efficiently before bringing in a receiver of any sort. Mammalian herbivores, such as wallabies and elephants, rely heavily on plant odours to find and choose their food (Stutz *et al.*, 2016; Schmitt *et al.*, 2018, 2020; McArthur *et al.*, 2019); and other herbivores, such as deer and antelope, likely use odour too. Defining the plant odour information they use will be crucial to understanding and manipulating their foraging for management and conservation goals.

Boundaries between cues and signals

Our results (Figs 3, 4) provide insight into possible differences in characteristics between cues and signals. The reduced set of

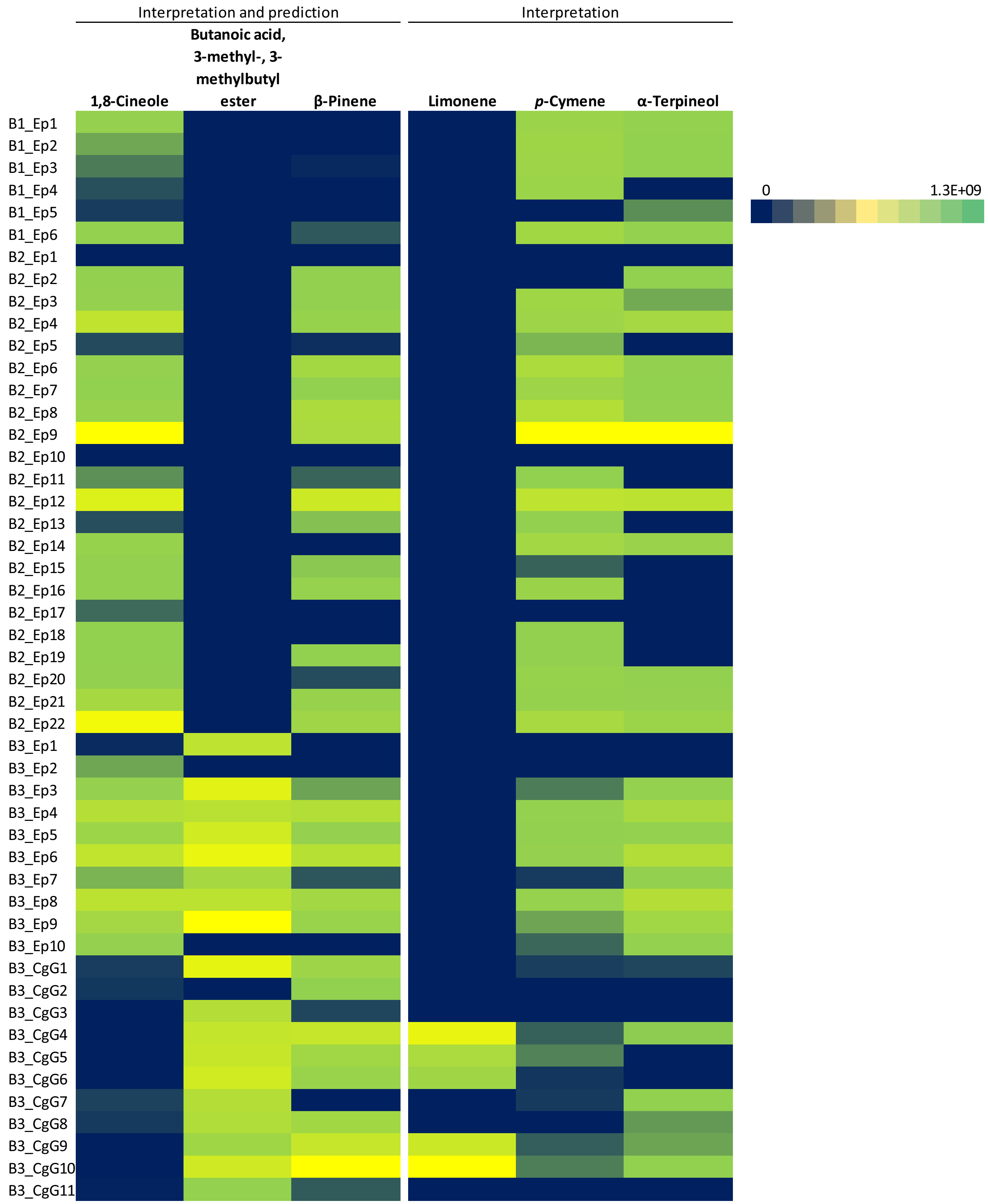

Fig. 7 Results from the random forest algorithm. The left-hand column lists each replicate taken for each batch (B) for our focal group (Ep, *Eucalyptus punctata*) and outgroup (Cg, *Corymbia gummifera*). The top row lists the six volatile organic compounds (VOCs) selected by the algorithm in the interpretation step. Of these six, only three remained after the prediction step (the three first columns from the left in the top row). The amount of emitted VOC (quantified as ion count) for each replicate is presented by colour from blue (zero) to green (highly abundant); see the key. For example, 1,8-cineole is mid to highly abundant in most *E. punctata* seedling replicates but is absent or near absent from *C. gummifera* seedling replicates.

potentially informative VOC pairs that we identified for undamaged *E. punctata* seedlings fell in a band between higher consistency/lower precision and vice versa, but no informative pairs were identified that were highly consistent and highly precise. This pattern matches the prediction from our conceptual framework that undamaged leaves, particularly of palatable plants, emit cues rather than signals (Fig. 2).

In future, our predictions about where cues and signals lie in consistency–precision reliability space can be tested more thoroughly by quantifying and comparing the distribution of potentially informative emissions (replicating the matrix in Fig. 3) from contrasting groups. Thus, in contrast to our results for undamaged leaves of a palatable plant, those of a highly unpalatable plant species may emit a signal as a form of olfactory aposematism (Lev-Yadun, 2021). Other examples predicted to have informative emissions reflecting cues vs signals respectively include the following: first, VOCs from plants with undamaged leaves vs plants with damaged leaves (and herbivore-induced plant VOCs (HIPVs, Turlings *et al.*, 1990)); second, flowers pollinated by wind vs those pollinated by animals; or third, unripe vs ripe fruits. Such empirical testing will confirm whether our conceptual framework captures the essential features of cues vs signals within the consistency–precision reliability space (Fig. 2), or whether it is less clear cut than we predict.

The ecological significance of producing different channels of information

Our results for *E. punctata*, using consistency of 0.67 and precision of 0.21 and considering accuracy (against the *C. gummifera* outgroup), detected three, or possibly four, potentially informative VOC pair combinations. These sets (as pairs or even as triplets) effectively comprise different channels of information, one or more of which could be used by a variety of receivers, such as pollinators, herbivores, or predators (Schiestl & Johnson, 2013).

From a receiver's perspective, the presence and use of different channels of information enable a backup in case one becomes unreliable. For example, an informative pair of VOCs emitted by an organism may be confounded by the same VOCs occurring in the background odour of natural environments (Schröder & Hilker, 2008; Riffell *et al.*, 2014). If one information channel is affected by background odour, those receivers making use of multiple channels may still be capable of detecting information. From an emitter's perspective, if interactions with other organisms are beneficial, then producing multiple channels of information as a backup becomes an advantage because it aids communication in noisy environments (Wilson *et al.*, 2015). Future studies could test whether channels of information are affected differently depending on the environment (e.g. closed forest vs woodland) and conditions (e.g. ambient temperature and light intensity) where they are released, and whether receivers can switch effectively among channels of information.

Our results for *E. punctata*, using consistency of 0.67 and precision of 0.21, also showed that 16% of all seedlings lacked any of the four informative VOC pairs identified. This suggests either that such seedlings escape detection by receivers or that information available for receivers lies in some other region of the consistency–precision reliability space (as in Fig. 4). On the flip side, 34% of all seedlings had one to three of the four informative pairs but only 50% had all four. This result may reflect redundancy, in which VOC pairs could be substituted for others without compromising the information; for examples, see Bruce & Pickett (2011).

Broader implications

The heuristic approach provides the basis for new insights into interactions mediated by olfactory information. These interactions include those with plants as emitters (e.g. plant–herbivore, plant–natural enemy, and plant–plant interactions), but more broadly any olfactory web, such as those interlinking predators, competitors, and prey (Banks *et al.*, 2016). We now offer the ability to delve into and characterize the information that underpins these interactions, to ask questions about how the information arises. Together with the conceptual framework (Fig. 2), our approach leads to novel ideas on how information variability could affect different interactions, which can, in turn, lead to the development of new management tools. For example, it could accelerate the development of effective biochemical sensors, such as e-noses, for pest management (Ivaskovic *et al.*, 2021).

Our heuristic method (and the algorithm we developed here) is equally applicable to identifying odour information emitted by any group. Thus, it should be useful for disciplines beyond ecology, from developing perfumes (e.g. by defining informative components of natural odours to create artificial ones) to training detection animals such as dogs and rats (MacDonald *et al.*, 2003) or elephants (Miller *et al.*, 2015). By identifying the most reliable VOCs of a focal source (e.g. prostate cancer; Elliker *et al.*, 2014) and using appropriate training (Fischer-Tenhagen *et al.*, 2017), detection rates by dogs may be improved.

Acknowledgements

This research was supported by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney. This research was supported by the Australian Research Council ARC-DP (DP190101441) to CM, PB, and MP; and a Murray Writing Up Support Grant 2021 (The University of Sydney) to CGO. Open access publishing facilitated by The University of Sydney, as part of the Wiley The University of Sydney agreement via the Council of Australian University Librarians.

Competing interests

None declared.

Author contributions

Concept and design of the research: CM, PB, CGO, MP and CP. Performance of the research/data collection: CGO, MP and

CP. Data analysis: CGO, MP, LM and CM. Data interpretation: CGO, CM, MP, CP, PB and LM. Writing the manuscript: CGO, CM and MP. Reviewing and editing the manuscript: CGO, CM, MP, CP, PB and LM.

ORCID

Peter B. Banks https://orcid.org/0000-0002-4340-6495
Clare McArthur https://orcid.org/0000-0002-7867-414X
Cristian Gabriel Orlando https://orcid.org/0000-0003-3937-423X
Malcolm Possell https://orcid.org/0000-0001-5376-8433
Catherine Price https://orcid.org/0000-0001-9631-2479

Data availability

Data supporting the findings of this study is included as Supporting Information.

References

Aartsma Y, Bianchi FJ, Werf W, Poelman EH, Dicke M. 2017. Herbivore-induced plant volatiles and tritrophic interactions across spatial scales. *New Phytologist* **216**: 1054–1063.

Ache BW, Young JM. 2005. Olfaction: diverse species, conserved principles. *Neuron* **48**: 417–430.

Aljbory Z, Chen MS. 2018. Indirect plant defense against insect herbivores: a review. *Insect Science* **25**: 2–23.

Auffarth B. 2013. Understanding smell – the olfactory stimulus problem. *Neuroscience & Biobehavioral Reviews* **37**: 1667–1679.

Banks PB, Daly A, Bytheway JP. 2016. Predator odours attract other predators, creating an olfactory web of information. *Biology Letters* **12**: 20151053.

Bedoya-Pérez MA, Isler I, Banks PB, McArthur C. 2014. Roles of the volatile terpene, 1,8-cineole, in plant–herbivore interactions: a foraging odor cue as well as a toxin? *Oecologia* **174**: 827–837.

Beyaert I, Hilker M. 2014. Plant odour plumes as mediators of plant–insect interactions. *Biological Reviews* **89**: 68–81.

Beyaert I, Wäschke N, Scholz A, Varama M, Reinecke A, Hilker M. 2010. Relevance of resource-indicating key volatiles and habitat odour for insect orientation. *Animal Behaviour* **79**: 1077–1086.

Birkett MA, Bruce TJ, Martin JL, Smart LE, Oakley J, Wadhams LJ. 2004. Responses of female orange wheat blossom midge, *Sitodiplosis mosellana*, to wheat panicle volatiles. *Journal of Chemical Ecology* **30**: 1319–1328.

Boland DJ, Brooker MIH, Chippendale G, Hall N, Hyland B, Johnston R, Kleinig D, McDonald M, Turner J. 2006. *Forest trees of Australia*. Collingwood, Vic., Australia: CSIRO.

Breiman L. 2001. Random Forests. *Machine Learning* **45**: 5–32.

Bruce TJ, Pickett JA. 2011. Perception of plant volatile blends by herbivorous insects–finding the right mix. *Phytochemistry* **72**: 1605–1611.

Bruce TJ, Wadhams LJ, Woodcock CM. 2005. Insect host location: a volatile situation. *Trends in Plant Science* **10**: 269–274.

Calogirou A, Kotzias D, Kettrup A. 1997. Product analysis of the gas-phase reaction of β-caryophyllene with ozone. *Atmospheric Environment* **31**: 283–285.

Campbell DR, Bischoff M, Lord JM, Robertson AW. 2010. Flower color influences insect visitation in alpine New Zealand. *Ecology* **91**: 2638–2649.

Conchou L, Anderson P, Birgersson G. 2017. Host plant species differentiation in a polyphagous moth: olfaction is enough. *Journal of Chemical Ecology* **43**: 794–805.

Cork SJ, Hume I, Dawson T. 1983. Digestion and metabolism of a natural foliar diet (*Eucalyptus punctata*) by an arboreal marsupial, the koala (*Phascolarctos cinereus*). *Journal of Comparative Physiology* **153**: 181–190.

Dearing MD, Foley WJ, McLean S. 2005. The influence of plant secondary metabolites on the nutritional ecology of herbivorous terrestrial vertebrates. *Annual Review of Ecology, Evolution, and Systematics* **36**: 169–189.

Dellacassa E, Menéndez P, Moyna P, Soler E. 1990. Chemical composition of *Eucalyptus* essential oils grown in Uruguay. *Flavour and Fragrance Journal* **5**: 91–95.

van den Dool H, Kratz PD. 1963. A generalization of the retention index system including linear temperature programmed gas–liquid partition chromatography. *Journal of Chromatography* **11**: 463–471.

Doty RL. 1986. Odor-guided behavior in mammals. *Experientia* **42**: 257–271.

Dunn JC. 1974. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics* **4**: 95–104.

Eisthen HL. 2002. Why are olfactory systems of different animals so similar? *Brain, Behavior and Evolution* **59**: 273–293.

Elaissi A, Salah KH, Mabrouk S, Larbi KM, Chemli R, Harzallah-Skhiri F. 2011. Antibacterial activity and chemical composition of 20 *Eucalyptus* species' essential oils. *Food Chemistry* **129**: 1427–1434.

Elliker KR, Sommerville BA, Broom DM, Neal DE, Armstrong S, Williams HC. 2014. Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from a double-blind, controlled trial of prostate cancer detection. *BMC Urology* **14**: 1–9.

Finnerty PB, McArthur C, Banks P, Price C, Shrader AM. 2022. The olfactory landscape concept: a key source of past, present, and future information driving animal movement and decision-making. *Bioscience* **72**: 745–752.

Finnerty PB, Stutz RS, Price CJ, Banks PB, McArthur C. 2017. Leaf odour cues enable non-random foraging by mammalian herbivores. *Journal of Animal Ecology* **86**: 1317–1328.

Fischer-Tenhagen C, Johnen D, Heuwieser W, Becker R, Schallschmidt K, Nehls I. 2017. Odor perception by dogs: evaluating two training approaches for odor learning of sniffer dogs. *Chemical Senses* **42**: 435–441.

Fraser AM, Mechaber WL, Hildebrand JG. 2003. Electroantennographic and behavioral responses of the sphinx moth *Manduca sexta* to host plant headspace volatiles. *Journal of Chemical Ecology* **29**: 1813–1833.

Gallahar N, Leigh K, Phalen D. 2021. Koala tree selection in a mixed-tenure landscape and post-fire implications. *Wildlife Research* **48**: 737–755.

Genuer R, Poggi J-M, Tuleau-Malot C. 2015. VSURF: an R package for variable selection using Random Forests. *The R Journal* **7**: 19–33.

Hiratani N, Latham PE. 2020. Rapid Bayesian learning in the mammalian olfactory system. *Nature Communications* **11**: 1–15.

Hoballah ME, Stuurman J, Turlings TCJ, Guerin PM, Connétable S, Kuhlemeier C. 2005. The composition and timing of flower odour emission by wild *Petunia axillaris* coincide with the antennal perception and nocturnal activity of the pollinator *Manduca sexta*. *Planta* **222**: 141–150.

Hodgkison R, Ayasse M, Kalko EKV, Häberlein C, Schulz S, Mustapha WAW, Zubaid A, Kunz TH. 2007. Chemical ecology of fruit bat foraging behavior in relation to the fruit odors of two species of paleotropical bat-dispersed figs (*Ficus hispida* and *Ficus scortechinii*). *Journal of Chemical Ecology* **33**: 2097–2110.

Hopfield J. 1999. Odor space and olfactory processing: collective algorithms and neural implementation. *Proceedings of the National Academy of Sciences, USA* **96**: 12506–12511.

Hu Z, Shen Y, Luo Y, Shen F, Gao H, Gao R. 2008. Aldehyde volatiles emitted in succession from mechanically damaged leaves of poplar cuttings. *Journal of Plant Biology* **51**: 269–275.

Inglis IR, Langton S, Forkman B, Lazarus J. 2001. An information primacy model of exploratory and foraging behaviour. *Animal Behaviour* **62**: 543–557.

Ivaskovic P, Ainseba BE, Nicolas Y, Toupance T, Tardy P, Thiéry D. 2021. Sensing of airborne infochemicals for green pest management: what is the challenge? *ACS Sensors* **6**: 3824–3840.

Jardine KJ, Monson RK, Abrell L, Saleska SR, Arneth A, Jardine A, Ishida FY, Serrano AMY, Artaxo P, Karl T. 2012. Within-plant isoprene oxidation confirmed by direct emissions of oxidation products methyl vinyl ketone and methacrolein. *Global Change Biology* **18**: 973–984.

Karban R, Shiojiri K, Huntzinger M, McCall AC. 2006. Damage-induced resistance in sagebrush: volatiles are key to intra-and interplant communication. *Ecology* **87**: 922–930.

Kessler A, Kalske A. 2018. Plant secondary metabolite diversity and species interactions. *Annual Review of Ecology, Evolution, and Systematics* **49**: 115–138.
Knight AL, Light DM. 2001. Attractants from Bartlett pear for codling moth, *Cydia pomonella* (L.), larvae. *Naturwissenschaften* **88**: 339–342.
Koops MA. 2004. Reliability and the value of information. *Animal Behaviour* **67**: 103–111.
Lawson CA, Seymour JR, Possell M, Suggett DJ, Raina J-B. 2020. The volatilomes of *Symbiodiniaceae*-associated bacteria are influenced by chemicals derived from their algal partner. *Frontiers in Marine Science* **7**: 106.
Lev-Yadun S. 2021. Avoiding rather than resisting herbivore attacks is often the first line of plant defence. *Biological Journal of the Linnean Society* **134**: 775–802.
Low PA, McArthur C, Fisher K, Hochuli DF. 2014. Elevated volatile concentrations in high-nutrient plants: do insect herbivores pay a high price for good food? *Ecological Entomology* **39**: 480–491.
MacDonald J, Lockwood J, McFee J, Altshuler T, Broach T, Carin L, Harmon R, Rappaport C, Scott W, Weaver R. 2003. *Landmine detection*. Arlington, VA, USA: RAND.
McArthur C, Finnerty PB, Schmitt MH, Shuttleworth A, Shrader AM. 2019. Plant volatiles are a salient cue for foraging mammals: elephants target preferred plants despite background plant odour. *Animal Behaviour* **155**: 199–216.
Miller AK, Hensman MC, Hensman S, Schultz K, Reid P, Shore M, Brown J, Furton KG, Lee S. 2015. African elephants (*Loxodonta africana*) can detect TNT using olfaction: implications for biosensor application. *Applied Animal Behaviour Science* **171**: 177–183.
Nicolè F, Guitton Y, Courtois EA, Moja S, Legendre L, Hossaert-McKey M. 2012. MSEASY: unsupervised and untargeted GC-MS data processing. *Bioinformatics* **28**: 2278–2280.
Ninkovic V, Markovic D, Rensing M. 2021. Plant volatiles as cues and signals in plant communication. *Plant, Cell & Environment* **44**: 1030–1043.
NIST. 2008. *The NIST mass spectral search program for the NIST/EPA/NIH mass spectral library*, v.2.0. Gaithersburg, MD, USA: National Institute of Science and Technology.
R Core Team. 2019. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing [WWW document] URL http://www.R-project.org.
Raguso RA. 2008. Wake up and smell the roses: the ecology and evolution of floral scent. *Annual Review of Ecology, Evolution, and Systematics* **39**: 549–569.
Ranganathan Y, Borges RM. 2010. Reducing the babel in plant volatile communication: using the forest to see the trees. *Plant Biology* **12**: 735–742.
Riffell JA, Lei H, Christensen TA, Hildebrand JG. 2009. Characterization and coding of behaviorally significant odor mixtures. *Current Biology* **19**: 335–340.
Riffell JA, Shlizerman E, Sanders E, Abrell L, Medina B, Hinterwirth AJ, Kutz JN. 2014. Flower discrimination by pollinators in a dynamic chemical environment. *Science* **344**: 1515–1518.
Rousseeuw PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**: 53–65.
RStudio. 2015. *RSTUDIO: integrated development for R, vol. 42.* Boston, MA, USA: RStudio [WWW document] URL http://www.rstudio.com
Saunders AS, Burgin S. 2001. Selective foliage foraging by Red Wattlebirds, *Anthochaera carunculata*, and noisy friarbirds, *Philemon corniculatus*. *Emu* **101**: 163–166.
Schiestl FP, Johnson SD. 2013. Pollinator-mediated evolution of floral signals. *Trends in Ecology & Evolution* **28**: 307–315.
Schmidt KA, Dall SR, Van Gils JA. 2010. The ecology of information: an overview on the ecological significance of making informed decisions. *Oikos* **119**: 304–316.
Schmitt MH, Shuttleworth A, Shrader AM, Ward D. 2020. The role of volatile plant secondary metabolites as pre-ingestive cues and potential toxins dictating diet selection by African elephants. *Oikos* **129**: 24–34.
Schmitt MH, Shuttleworth A, Ward D, Shrader AM. 2018. African elephants use plant odours to make foraging decisions across multiple spatial scales. *Animal Behaviour* **141**: 17–27.
Schröder R, Hilker M. 2008. The relevance of background odor in resource location by insects: a behavioral approach. *Bioscience* **58**: 308–316.
Southwell IA. 1973. Variation in the leaf oil of *Eucalyptus punctata*. *Phytochemistry* **12**: 1341–1343.
Spiegel O, Crofoot MC. 2016. The feedback between where we go and what we know – information shapes movement, but movement also impacts information acquisition. *Current Opinion in Behavioral Sciences* **12**: 90–96.
Steiger S, Schmitt T, Schaefer HM. 2010. The origin and dynamic evolution of chemical information transfer. *Proceedings of the Royal Society B: Biological Sciences* **278**: 970–979.
Steinbauer M. 1996. A note on manna feeding by ants (Hymenoptera: *Formicidae*). *Journal of Natural History* **30**: 1185–1192.
Stutz RS, Banks PB, Proschogo N, McArthur C. 2016. Follow your nose: leaf odour as an important foraging cue for mammalian herbivores. *Oecologia* **182**: 643–651.
Tromelin A. 2016. Odour perception: a review of an intricate signalling pathway. *Flavour and Fragrance Journal* **31**: 107–119.
Turlings TC, Tumlinson JH, Lewis WJ. 1990. Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science* **250**: 1251–1253.
Visser J, Avé D. 1978. General green leaf volatiles in the olfactory orientation of the Colorado beetle, *Leptinotarsa decemlineata*. *Entomologia Experimentalis et Applicata* **24**: 738–749.
Vos M, Vet LE, Wäckers FL, Middelburg JJ, Van Der Putten WH, Mooij WM, Heip CH, Van Donk E. 2006. Infochemicals structure marine, terrestrial and freshwater food webs: implications for ecological informatics. *Ecological Informatics* **1**: 23–32.
Webster B, Bruce T, Dufour S, Birkemeyer C, Birkett M, Hardie J, Pickett J. 2008. Identification of volatile compounds used in host location by the black bean aphid, *Aphis fabae*. *Journal of Chemical Ecology* **34**: 1153–1161.
Webster B, Bruce T, Pickett J, Hardie J. 2010a. Volatiles functioning as host cues in a blend become nonhost cues when presented alone to the black bean aphid. *Animal Behaviour* **79**: 451–457.
Webster B, Gezan S, Bruce T, Hardie J, Pickett J. 2010b. Between plant and diurnal variation in quantities and ratios of volatile compounds emitted by *Vicia faba* plants. *Phytochemistry* **71**: 81–89.
Wiggins NL, McArthur C, Davies NW. 2006. Diet switching in a generalist mammalian folivore: fundamental to maximising intake. *Oecologia* **147**: 650–657.
Wilson JK, Kessler A, Woods HA. 2015. Noisy communication via airborne infochemicals. *Bioscience* **65**: 667–677.
Zini CA, Augusto F, Christensen E, Caramão EB, Pawliszyn J. 2002. SPME applied to the study of volatile organic compounds emitted by three species of *Eucalyptus in situ*. *Journal of Agricultural and Food Chemistry* **50**: 7199–7205.
Zu P, Boege K, Del-Val E, Schuman MC, Stevenson PC, Zaldivar-Riverón A, Saavedra S. 2020. Information arms race explains plant–herbivore chemical communication in ecological communities. *Science* **368**: 1377–1381.
Zuberbühler K, Noë R, Seyfarth RM. 1997. Diana monkey long-distance calls: messages for conspecifics and predators. *Animal Behaviour* **53**: 589–604.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 Relative abundance of the volatile organic compounds identified for *Eucalyptus punctata*.

Dataset S2 Relative abundance of the volatile organic compounds identified for *Corymbia gummifera*.

Methods S1 Supplementary information for the materials and methods used in this study.

Notes S1 R code created to run both the 'Odour Information Exploration' and the 'Odour Information Definition' algorithms for the new heuristic method proposed in this study.

Notes S2 R code used to run the random forest algorithm.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.