How can population-level chemical diversity evolve through individual-based selection? An approach to resolve Crozier's paradox

Marti Wittke & Florian Menzel

Institute of Organismic and Molecular Evolution, Johannes Gutenberg Universität Mainz, Germany

Introduction

Trait variation among conspecific individuals is ubiquitous, and has important effects on biotic communities and ecosystems (Bolnick et al., 2011; Schuman et al., 2016). Especially for communication, trait diversity concerning chemical or optical cues is the prerequisite for any form of recognition, be it in individualised groups (Gilfillan et al., 2016; Tibbetts & Dale, 2007) or in social insects, where individuals recognise group membership rather than individual entities. In eusocial insects in particular, recognition is crucial to maintain the integrity of the colony. Here, nestmate recognition ensures that behavioural altruism is directed only towards members of the own colony (Leonhardt et al., 2016; Vandermeer & Morel, 1998). In social insect colonies, like ants, bees or wasps, individuals discriminate between nestmates and non-nestmates, and decide whether or not an encountered individual is granted access to the nest and its resources. Nestmate recognition is mediated by cuticular hydrocarbons (CHCs), which form complex blends of up to 150 different compounds on a single individual (Sprenger & Menzel, 2020). There is strong selection to avoid that the nest is exploited by alien, unrelated individuals. Cheaters, on the other hand, often mimic others so that they are not recognised, i.e. they are selected to be as similar as possible to a potential 'victim' colony. Such chemical mimicry has frequently evolved in socially parasitic ants, but also in myrmecophilic insects such as staphylinid beetles or lycaenid butterflies (Guillem et al., 2014; Schlick-Steiner et al., 2004).

However, the evolution of recognition cues raises a theoretical problem concerning the direction of selection pressures: if there is high aggression against individuals from foreign colonies, then each colony should benefit from resembling the others (Crozier, 1986). Aggression between opponents often increases with increasing chemical distance (Kidokoro-Kobayashi et al., 2012; Wittke et al., 2022). Therefore, each individual should elicit less aggression if its chemical signature differs less from its opponent, and be selected to be as similar to the others as possible. However, this would result in a homogenous odour across the entire population of colonies – which would thwart all possibilities to actually perform nestmate recognition, since discrimination only works if groups differ from each other. This reasoning – that, in theory, there should be selection against being different from others - has been termed *Crozier's paradox.* But then, why does cue diversity not decline in natural populations, and how can diversity evolve at all? Crozier's paradox has the additional interesting twist of how individual-based selection (survival and fitness of individual colonies) can result in a population-level trait such as cue diversity across multiple colonies. Note that, in eusocial insects, selection acts on the colony level, since this is the reproductive unit.

A previous approach to solve Crozier's paradox used an agent-based model to show that cue diversity can be maintained if there is disassortative mating (Holman et al., 2013). Indeed, CHCs were shown to be sexually selected in solitary insects (Steiger et al., 2013; Steiger & Stökl, 2014). In social insects, CHCs mediate nestmate recognition, hence recognition cues may be under sexual selection as well. However, this idea rests on the assumption that insects mate disassortatively. This has been shown in solitary insects (Lihoreau & Rivault, 2009; Thomas & Simmons, 2011) but evidence for eusocial insects is scarce. Moreover, intra-sexual and inter-sexual selection in social insects seems to be weak, since the mating system in social

Hymenoptera leaves little room for female choice. Thus, variation in male reproductive success (number of matings) is far lower than in other organisms and males cannot monopolise multiple females due to once-in-a-lifetime partner selection (via mating and sperm storage) (Boomsma et al., 2005; Strassmann, 2001). Hence, it is unclear whether the disassortative mating required for previous models actually holds for many social insects.

A further, often-debated question concerns how exactly the recognition system determines the chemical distance of its opponent to an internal, neuronal template (which is the first step of the recognition process). Firstly, it might use the overall similarity, averaging chemical distance over all chemical compounds present. This has been suggested as the *Gestalt* model (Crozier & Dix, 1979; Vandermeer et al., 1998). Alternatively, the ant might only consider whether the opponent bears compounds that are not part of its internal template (*'undesired-absent'* or *U-absent)* or whether it bears all compounds that are part of the template (*'desired-present'* or *D-present*) (d'Ettorre & Lenoir, 2010). These models imply that the ant ignores any missing cues that are present in its template in the first case, or any additional cues not present in its template in the second one. Evidence supports *U-absent* rather than *D-present* (Nehring et al., 2016; van Zweden & d'Ettorre, 2010).

Here, we used an agent-based model to study how cue diversity can evolve despite Crozier's paradox, and compare the three recognition models mentioned above. Our model does not require disassortative mating. Instead, we simulate a population of social insect colonies that can invade and exploit each other. Each individual colony can forage on its own, but can also invade other colonies to exploit their resources; thus parasite and host are the same species, and each colony can adopt both roles simultaneously. Only the most successful colonies pass on their profiles to the next generation. Using this model, we studied how population-level cue diversity evolved over time. Furthermore, we assessed which recognition model gave the most realistic results, and how cue evolution was affected by different levels of signal production costs, selection strength and rate of cue evolution (evolvability).

Materials & Methods

The model

The model was developed in NetLogo version 6.2.2 (Wilensky, 1999). The simulated world contains 50 ant colonies with 10 workers each. The number of workers stays constant unless the colony dies (see below). At the beginning of each timestep, 300 food resources are created, each consisting of a single unit of food.

Recognition signal

Each colony has with a chemical profile which serves as recognition signal. In social insects, recognition is usually mediated by cuticular hydrocarbons (CHCs), which form complex mixtures that encode, among others, nestmate membership (Sprenger & Menzel, 2020; Sturgis & Gordon, 2012). In our model, the signal consists of 10 compounds, each with an abundance of zero to infinite. For a colony *r*, the abundances of the ten compounds are denoted as $c_{r,1}$ to $c_{r,10}$ and can vary independently from each other. Total cue abundance of colony *r* is calculated as

$$C_r = \sum_{n=1}^{10} c_{r,n}.$$

In ant CHCs, not all CHCs vary independently (S. Martin & Drijfhout, 2009; S. J. Martin & Drijfhout, 2009). This may be due to physical, but also to biosynthetic constraints (F. Menzel et al., 2017). Hence, the 'compounds' of our model should not necessarily be seen as real CHCs, but rather as entities (e.g. groups of compounds) that vary independently from each other. For example, they might be viewed as homologous series of CHCs, such as the sum of

15-MeC25, 15-MeC27 and 15-MeC29, or as a principal component axis along which sets of hydrocarbons co-vary and which are by definition independent from each other. Initial abundance for each colony was generated by drawing from an exponential distribution with an expected value of 10.

Individual ants have a profile that varies around the above-described colony-specific profile. The profile of each individual was generated by drawing the abundances of each compound from normal distributions with (mutation strength, see below).

Foraging and colony-level resource consumption

In each timestep (henceforth, "tick"), each individual ant tries to obtain food.Each colony and food-source (initial value at each tick: 300) had an equal probability to be targeted when collecting food Thus, the initial probability to raid a colony instead of collecting food is 49/350, but increases as the food sources get depleted in the course of each tick. If the individual collects a food source, it will gain one unit of food, leaving the food source depleted. Raiding involves trying to enter the other colony, where a recognition process (see below) between the intruding ant and the resident colony takes place. This process decides whether the intruder is rejected and returns to its home colony, or whether it is allowed to the colony, steals one unit of food (and the raided colony loses 1 unit of food) and then returns to the home colony. The order of individuals performing a behaviour is randomized at each tick, thus simulating a simultaneous activity of all colonies, i.e. in no tick does any colony have an advantage or disadvantage by the order they are acting in. Any food source that was not collected in the previous time step is removed, which simulates other animals foraging and food perishing, and new food sources are generated, i.e. food cannot accumulate.

Each colony r has a resource stock at time t of $R_{r,t}$ with the initial value $R_{r,1} = 25$. Metabolic costs of the colony's ants are subtracted after each tick, such that

$$R_{r,t+1} = R_{r,t} - \frac{f * C_r}{1000}$$

Where C_r is the total cue abundance and f is the cost of CHC biosynthesis per unit CHCs. This formula was chosen so that costs increased with the overall quantity of cues produced. Colonies died if they had no more than 0 units at the end of each tick, being replaced only at the end of the generation.

Colony invasion and the nestmate recognition process

Every time an individual ant attempts to invade another colony, a nestmate recognition process determines whether it will be successful. One of the ants of the invaded colony ('resident') perceives the intruder's profile and calculates the chemical distance between the intruder and the average of the profiles of the resident colony. This distance ranges from 0 to 1 and determines the chance of raiding success, which is central to this model. The probability that an intruder is allowed to the colony is equal to this distance. In case of rejection, it immediately returns to its home colony without food. If it is accepted, the target colony loses one unit of food to the intruder, which then returns to its home colony.

The same process takes place when, after foraging, an ant returns to its home colony. Only if it is considered a nestmate, its food is added to the colony stock. If (erroneously) not, the food is discarded but the ant can keep foraging in the next tick and re-try to enter its home colony. This was implemented so that ants would not only have to recognise and reject non-nestmates, but also to recognise (and allow) nestmates.

For the recognition process, we considered three recognition models: *Desirable-present* (*D-present*), *undesirable-absent (U-absent)* and *overall similarity*. These models determine the way the chemical distance between opponent and target colony is calculated. *Overall similarity* is similar to the conjectured recognition in the *Gestalt* model (Crozier & Dix, 1979; Vandermeer

Kommentiert [FM1]: Ich habe das jetzt so formuliert, aber es ist noch etwas ungenau, finde ich. Du hast also diese Distanz, und ziehst dann eine Zahl aus einer Gleichverteilung. Wenn diese Zahl höher ist als die Distanz, ist der Angreifer erfolgreich, wenn nicht, dann nicht. Was machst du, wenn beide Zahlen gleich hoch sind? Wie würdest du das im Text formulieren?

Kommentiert [2]: *Antwort auf Menzel, Dr. Florian (03.12.2023, 15:44): "..."*
Nein, eine Gleichverteilung (d.h. jede Zahl ist gleich wahrscheinlich), entsprechend ist der Abstand zwischen Angreifer/Verteidiger gleich der Wahrscheinlichkeit, als Dieb erkannt zu werden.
D.h. bei maximalem Abstand (=1) könnte nie gestohlen werden. Bzw. Umgekehrt ist stehlen immer erfolgreich wenn der Abstand = 0 ist (war so richtig im Text).

Kommentiert [FM3]: Verstehe. Ich hatte mit "Laplace-Verteilung" eine Gleichverteilung gemeint, war irrtümlich der Meinung, die hieße so.
Anyway - muss die Zufallszahl größer oder größer gleich Distanz sein, damit es erfolgreich ist?

Kommentiert [MW4]: Hatte ich auch mal gedacht (weil ein Laplace-Experiment eins mit gleichen Wahrscheinlichkeiten für jedes Ergebnis ist). Passiert, wenn man alles nach den selben paar Mathematikern benennt

Technisch betrachtet größer-gleich (im Code hat man einen Misserfolg wenn die Zahl kleiner ist, und einen Erfolg dann umgekehrt wenn größer-gleich).

Mathematisch macht das lustigerweise keinen Unterschied. In einer stetigen Zufallsverteilung (wie eine gleichverteilung in auf den reelen Zahlen) ist die Chance eine Zahl *genau* zu ziehen immer 0, weil es auch im Bereich 0-1 unendlich viele Zahlen gibt (die sich alle irgendwo in einer Nachkommastelle unterscheiden). Der Fall das die Zufallszahl gleich dem Abstand ist taucht also mathematisch nicht auf, sondern jede gezogene Zahl ist immer entweder größer oder kleiner, also mathematisch ist P(x<y) das gleiche wie P(x<=y).

Das aber mehr als kleiner Ausflug in die Statistik, spielt eigentlich keine Rolle^^.

& Morel, 1998) and is calculated as the Bray-Curtis distance between the profile of the encountered (thieving) ant and the colony profile of the colony it intends to enter. The distance between the resident colony *r* and the intruder *i*, according to the recognition model, is calculated as

$$BC_{r,i}^{G} = 1 - \frac{2\sum_{k=1}^{n}\min\left(c_{k,r}; c_{k,i}\right)}{\sum_{k=1}^{n} c_{k,r} + c_{k,i}}$$

with $c_{n,r}$ and $c_{n,i}$ as the abundance of cue *k* in the resident colony and the intruding individual, respectively. In our simulation we simulated profiles as having 10 components, thus n = 10.

The *U-absent* model focuses on 'unknown' compounds that are not present or less abundant in the ant which perceives an the opponent's (the intruder's) cue profile. Thus, it ignores compounds that are present in the resident, but absent or less abundant in the intruder ($c_i \leq c_r$). To implement this, we identified those compounds and set them to equal values in resident and intruder (here: c_i), and calculated the Bray-Curtis distance as above. This way, compounds less abundant in the intruder did not contribute to the overall Bray-Curtis distance. In the *D-present* model, we placed an emphasis on whether the opponent ant possessed all compounds of the resident ant. Hence, we identified compounds that were less abundant in the resident, and set them to equal values in resident and intruder (c_r). This way, compounds that were less abundant in the resident were ignored. The formulae for Bray-Curtis distance under the *U-absent* and the *D-present* model could hence also be written as: *(Florian's comment: these formulae are new and may be not the final notation)*

$$BC_{r,i}^{DP} = 1 - \frac{2\sum_{k=1}^{n}\min\left(c_{k,r}; c_{k,i}\right), else}{\sum_{k=1}^{n}\begin{cases} 2c_{k,r}\ if\ c_{k,i} > c_{k,r} \\ c_{k,r} + c_{k,i}\ otherwise \end{cases}}$$

$$BC_{r,i}^{UA} = 1 - \frac{2\sum_{k=1}^{n}\min\left(c_{k,r}; c_{k,i}\right), else}{\sum_{k=1}^{n}\begin{cases} 2c_{k,i}, wenn\ c_{ki} < c_{kr} \\ c_{k,r} + c_{k,i}, sonst \end{cases}}$$

Reproduction and selection

The simulated colonies reproduce and die seasonally. After a generation time of 100 ticks, some colonies reproduce, whereas others die. In principle, however, a colony can live forever. At the end of each generation, a defined percentage ('selection strength', *s*) of the colonies dies. The remaining ones produce offspring colonies, with the number of offspring colonies being proportional to their amount of collected food relative to the total amount of food available in all living colonies. The number of offspring colonies is adjusted so that each generation starts with the same number of colonies. This includes replacing colonies that had died from starvation.

Offspring colonies inherit their profiles from their parent colony, but with random variation ('mutation strength', *m*). If the parent colony has the profile $c_{r,1}\ldots c_{r,10}$, a daughter colony will have cue abundances drawn from a normal distribution with the means $c_{r,1}\ldots c_{r,10}$ and the standard deviation *m*. *m* is to reflect the influence of mutations, but also environental variation, on CHC variation, and can be seen as a proxy of evolutionary malleability of a CHC profile (as opposed to its phylogenetic conservation). Compound abundance has no maximum, but is capped at a minimum of 0.

Experiments performed and data recorded

Our main question was how profile diversity evolves, and whether it is maintained over the course of time. Therefore, we performed a simulation with the model as described above (parameter settings $s = 0.4$, $f = 40$ and $m = 5$). We varied the three recognition models (*overall similarity, D-present, U-absent*) and performed a control experiment in which the selection of colonies that reproduced or died at the end of a generation was done randomly, and not based on the amount of food collected. Each experiment ran for 2000 ticks (i.e. 20 generations) and was replicated 100 times. At the end of each simulation, we calculated population-level cue diversity as the average of all pairwise Bray-Curtis distances between all colonies. Secondly, we recorded the mean absolute cue abundance C_r across all colonies. This was to assess how cue abundance evolves, and how their evolution is shaped by recognition models and profile costs.

To assess the robustness of our results and assess the effect of each individual parameter, we conducted simulations where we varied one of the main parameters. We varied (1) selection strength s (s = 0.1, 0.4 or 0.8), (2) CHC costs (f = 20, 40 or 80), and (3) mutation strength (m = 1, 5 or 20). In each experiment, the other parameters were set to the default values (s = 0.4, f = 40, m = 5).

We conducted 100 replicate simulations for each parameter combination. In each replicate, we recorded the above-mentioned two data points at the end of each generation (i.e. 20 sets of 2 values per replicate). For each of the three experiments, we conducted 1200 simulations (3 parameter levels x (3 recognition models + control) x 100 replicates). The effect of each parameter was assessed using linear models. In separate models, we used final cue diversity and final cue abundance as response variables, the explanatory variables being *recognition model* (*U-absent, overall similarity, D-present, control)* and one of the parameters *profile cost, selection strength*, and *evolvability*. Interactions between the parameter in question and *recognition model* were allowed. Effect sizes for single recognition models or parameter levels (t values) were obtained from model summaries. Furthermore, we tested whether cue diversity and cue abundance (respectively) changed over time, by comparing initial (tick 99) and final (tick 1999) values using paired t-tests (Tables S1, S2).

Put somewhere:

While it is hypothesised that the costs to produce cuticular hydrocarbons less than 1% of the basic metabolism (Dirks & Federle, 2011), the precise costs are unknown.

Results

Main experiment

Generally, cue diversity increased over time for all recognition models, but not in the control experiment (Fig. 1a). The final cue diversity differed between all four recognition models (F_3 = 885.3, $p < 0.0001$). It was highest for the *U-absent* model, followed by *overall similarity* and then *D-present*, which was still higher than the control. In the control experiment, diversity significantly declined over time (Table S1).

The absolute CHC quantity declined for all three recognition models compared to the start values, and only increased in the control setting (Fig. 1b, Table S1). All four models differed significantly from each other (F_3 = 3657.1, $p < 0.0001$), with the highest CHC quantity in the control setting, followed by *D-present*, *overall similarity*, and the lowest quantity in *U-absent*.

Selection strength

Increasing selection strength led to even higher cue diversity for the *U-absent* and the *overall similarity* model. For the *D-present* model, cue diversity peaked at intermediate selection strength, while it declined with selection strength for the control experiment. Overall, selection strength thus had a significant effect that depended on the recognition model (recognition model: $F_3 = 1624.0$, $p < 0.0001$, selection strength: $F_1 = 3.9$, $p = 0.049$, interaction: $F_3 = 343.8$, $p < 0.0001$) (Fig. 2). It was positive for *U-absent* and *overall similarity* (both $t > 13$, $p < 0.0001$), non-significant for *D-present* ($t = 0.24$, $p = 0.81$) and negative for *control* ($t = -25.0$, $p < 0.0001$). However, irrespective of selection strength, cue diversity was always highest for the *U-absent* model, followed by *overall similarity*, then *D-present* and finally *control*. Compared to the initial settings, cue diversity always increased and cue abundance always decreased, except for the *control* model. Effect sizes of *t* tests were highest for *U-absent*, followed by *overall similarity* and *D-present* (Table S2).

Cue abundances showed no clear trend with selection strength, increasing only for the *control* model (recognition model: $F_3 = 5017.3$, $p < 0.0001$, selection strength: $F_1 = 196.0$, $p < 0.0001$, interaction: $F_3 = 273.6$, $p < 0.0001$). As in the main model, they were lowest for *U-absent*, followed by *overall similarity, D-present* and *control* (Fig. 2).

Costs of CHC biosynthesis

Higher profile costs often led to an increase of cue diversity (Fig. 3). However, this effect was weakest in *U-absent* ($t = 6.0$, $p < 0.0001$), moderate in *overall similarity* ($t = 17.5$, $p < 0.0001$), and high in *D-present* ($t = 37.4$, $p < 0.0001$). No effect of profile cost was found in the *control* assay ($t = -0.4$, $p = 0.66$) (recognition model: $F_3 = 2651.8$, $p < 0.0001$, profile cost: $F_1 = 913.2$, $p < 0.0001$, interaction: $F_3 = 276.6$, $p < 0.0001$).

As expected, higher profile costs led to a decrease in cue abundance. Again, this effect was weakest in *U-absent* ($t = -3.8$, $p < 0.0001$), moderate in *overall similarity* ($t = -12.0$, $p < 0.0001$), and high in *D-present* ($t = -48.2$, $p < 0.0001$). Thus, low profile costs led to very high cue quantities in the *D-present* model, but less so in the other two models. No effect of profile cost was found in the *control* assay ($t = 1.2$, $p = 0.22$) (recognition model: $F_3 = 5569.2$, $p < 0.0001$, profile cost: $F_1 = 984.7$, $p < 0.0001$, interaction: $F_3 = 501.4$, $p < 0.0001$). For all three levels of profile cost, cue abundances were highest in the control, followed by the *D-present* model (Fig. 3). Comparisons to the initial values revealed increases in cue diversity and decreases in cue abundance for all recognition models except for the control (Table S2).

Evolvability

Higher evolvability generally led to higher cue diversity ($F_1 = 2353.3$, $p < 0.0001$) (Fig. 4). However, the effect size differed between models (interaction: $F_3 = 124.8$, $p < 0.0001$; main effect of model: $F_3 = 169.2$, $p < 0.0001$). Effect sizes of evolvability were only similar for *overall similarity* and *U-absent*, but different for all other pairs of models. Interestingly though, cue diversity increased over time only for evolvability 5 and 20, but not for evolvability 1 (Table S2). For evolvability 1, cue diversity strongly decreased over time, with strongest decrease in *U-absent* and weakest in *D-present* (Table S2, Fig. S3).

Like cue diversity, cue abundance also increased with evovability ($F_1 = 8442.8$, $p < 0.0001$). Again, effect sizes differed between models (interaction: $F_3 = 4537.8$, $p < 0.0001$; main effect of model: $F_3 = 8490.4$, $p < 0.0001$). Here, effect sizes differed significantly between each pair of models.

Discussion

Crozier's paradox states that in theory, population-level diversity of recognition cues should erode quickly, because individuals are attacked less if they are similar to others, and hence benefit from being similar to the population average. However, this would render recognition impossible, which relies on inter-individual cue differences. Here, we investigated a potential solution to this scenario, and showed that cue diversity on population level can evolve based on individual-level selection. The main new angle of our approach compared to previous models (Holman et al., 2013; Rousset & Roze, 2007) is that cue perception is separated from the actual cue profile. Therefore, two individuals with different cue profiles need not necessarily consider each other as different. Cues and cue recognition mechanisms evolve independently, and the same cues can be interpreted differently by different individuals. Under this scenario, an individual may profit from being different from the population average, because it may be more efficient in recognising and fending off intruders.

The separation of cue perception and the profile itself is realistic: cue production and cue perception, i.e. 'sender' and 'receiver' side, are separate pathways that evolve separately albeit they exert selection onto each other (Leonhardt et al., 2016). Cue perception involves the production of odorant receptors as well as the neuronal template to which the perceived profiles are compared. In social insects, odorant receptors diversified drastically, which underlines their importance for social interactions and for recognition (d'Ettorre et al., 2017; McKenzie et al., 2016; Mier et al., 2022; Pask et al., 2017). Furthermore, odorant receptor expression can evolve quickly and even change during an insect's lifetime, e.g. when they switch behavioural castes (Caminer et al., 2023). In ants for example, social parasites lost odorant receptors, exemplifying that odorant receptor suites can evolve rapidly (Jongepier et al., 2022).

Given that in *D-present* and *U-absent* not all CHCs are perceived, the perceived chemical distances between two individuals may be different, depending on the recognition model they use, their number of odorant receptors, but also depending on their own profile. Thus, an individuals A may regard individual B as different, but B may regard A as similar. This concept was not considered in previous models, but is implemented in our model through the separation of cue profile and cue perception.

Our model assumes a scenario of intraspecific parasitism. Colonies can forage on their own, but also raid food sources of other colonies, and thus are host and parasite simultaneously. Such a scenario has been shown e.g. in the neotropical *Ectatomma ruidum* (Formicidae: Ectatomminae), where intraspecific food robbing (cleptoparasitism) occurs and can reduce colony productivity (Jandt et al., 2015). Next to resources being thought of as food resources, it is also conceivable to think about them as workforce (larvae or pupae) that is being robbed – which would represent social parasitism. Indeed, intraspecific social parasitism is common in many ant species, especially as temporary social parasitism during colony foundation (Buschinger, 2009; Seifert, 2018). Also in established colonies, older and larger colonies often raid brood from younger colonies to enlarge their own size, which has been shown for Myrmicine and Ponerine ants (Paul et al., 2016; Stamps & Vinson, 1991). Social parasite species are often the closest relatives to their host species (Buschinger, 2009). This suggests that interspecific parasitism evolved from intraspecific parasitism, which further justifies the scenario we assumed. Parasitic interactions can shape the evolution of host diversity and host recognition systems. Parasitism can favour host diversification via exerting negative frequency-dependent selection (Hamilton, 1982; van Valen, 1973). This has also been shown for recognition-relevant traits, e.g. for cuckoos (Kilner & Langmore, 2011; Øien et al., 1995) and for ants where parasitised populations carry higher cue diversity than non-parasitised ones (Jongepier & Foitzik, 2016). This matches our results, where intraspecific parasitism led to higher cue diversity.

Our model is parsimonious in that it does not make any assumptions concerning disassortative mating or sexual selection. While one study showed that ants can mate disassortatively (i.e. they prefer non-nestmates) (Oppelt et al., 2008), further evidence for sexual selection is generally low in ants, and there seem to be few opportunities for males or females to actually choose between different mates (Heinze pers. Com.; (Boomsma et al., 2005).

Parasitism can favour host diversification via exerting negative frequency-dependent selection (Hamilton, 1982; van Valen, 1973). This has also been shown for recognition-relevant traits, e.g. for cuckoos (Kilner & Langmore, 2011; Øien et al., 1995) and for ants where parasitised populations carry higher cue diversity than non-parasitised ones (Jongepier & Foitzik, 2016). This matches our results that intraspecific parasitism can lead to higher cue diversity (and in theory would also promote the evolution of more sophisticated/sensitive recognition systems).

Evolutionary trajectory of diversity in the different recognition models

In our simulations, cue diversity increased over time. This was true for a range of values concerning evolvability, production costs and selection strength, thus showing that this result is robust to a wide variety of parameter combinations. Notably, *overall similarity* and *U-absent* always led to higher cue diversity, and at the same time lower cue abundance, compared to *D-present*. This is confirmed by empirical evidence, which supports *U-absent* rather than *D-present* to be the more likely recognition mechanism (van Zweden & d'Ettorre, 2010) (Nehring et al. 2016). Also from a theoretical point of view, the *D-present* model seems disadvantageous, because it is easier to trick by simply producing high amounts of recognition cues of a range of potential hosts. This also explains why the absolute cue quantities strongly increase over time in *D-present*, but less so in the other two models.

Interestingly, cue diversity only decreased over time for very low evolvability (i.e. low mutation rate), and here especially in the *U-absent* model (Table S2). This may actually be a case where Crozier's paradox applies – here, diversity declined strongly, possibly due to selection against being different (Crozier, 1986). This selection was not counteracted, because mutations that cause cue diversification were rare (low evolvability). In the *U-absent* model, many of the initial lineages went extinct, until only 1-3 of them remained. This can be seen by the strong decrease of cue diversity after ca. 5 generations (Fig S3). This is probably due to selection to be close to the average, which led to the extinction of those lineages that happened to differ most from the population average. Only after this extinction, there was a weak increase (Fig. S2). This 'extinction dip' was considerably weaker in the *overall similarity* model and did not occur at all in the *D-present* model. We hypothesise that in *D-present,* considerably more original lineages survived because colonies could escape the selection to the average by expressing a higher overall cue abundance, such that they would fulfil the 'being similar' criterion for multiple templates.

The *U-absent* model selects for components being less abundant compared to the opponent's template, which also means low total cue abundances. The theoretically optimal cue profile is thus one where all components are zero. Notably, this strategy ('chemical insignificance') is used by some social parasites and myrmecophiles (Lenoir et al., 2001; Lorenzi & d'Ettorre, 2020), but did not evolve here. In contrast, the *D-present* model selects for high abundances, because the opponent is considered similar if it expresses as many components of the own profile as possible.

Effects of profile costs

Interestingly, high profile costs always led to an increase in cue diversity. This effect was strongest in *D-present*: Here, low profile costs led to an enormous increase in cue production

(cue abundance), at the cost of cue diversity. At high overall abundances, a defined change in cue composition has a lower relative effect, which may explain why cue diversity was lower at high cue abundances. Hence, profile costs actually promote the evolution of cue diversity. This effect was always highest in *U-absent*, followed by *overall similarity* and then *D-present.* Thus cue diversity can evolve irrespective of the costs of cue biosynthesis, which might facilitate further research on cue evolution given that the costs of CHC biosynthesis have been notoriously difficult to quantify,

Conclusions

We showed that population-level cue diversity can evolve in a scenario of intraspecific parasitism if cues and cue perception are considered separately. The effect is robust for various parameter values. Only for very low evolvability, Crozier's paradox actually applies, leading to a decrease in cue diversity. Interestingly, even high cue production costs do not hinder the evolution of cue diversity.

Future studies could focus on the question how cue perception and the precision of the recognition template evolve. Firstly, the information gained from perceiving a cue profile may not always be sufficient to decide whether the opponent is a nestmate or not. There is evidence of such recognition uncertainty, e.g. in ants with very long-chain CHCs, which may be harder to perceive (Florian Menzel et al., 2013; Florian Menzel, Linsenmair, & Blüthgen, 2008). The notion that a clear nestmate discrimination requires a certain cue quantity is corroborated by studies showing that increasing CHC quantities also increase aggressive responses, but only for non-nestmate CHCs (Cini et al., 2009; Torres et al., 2007), and shows why chemical insignificance is a common strategy of myrmecophilic ants (Lenoir et al., 2001; Lorenzi & d'Ettorre, 2020). Too low amounts of recognition cues – or too small aberrations from the own profile may also explain lack of aggression in intercolonial encounters (Foitzik et al., 2007; Steiner et al., 2007). Here, different costs of type I and type II errors (acceptance of foreigners vs. rejection of nestmates) may affect how the template and behavioural responses evolve. Modelling the evolution of template precision (e.g. number and richness of odorant receptors) may further enable us to understand the evolution of nestmate recognition, and especially to understand how supercolonies evolve, and whether they are evolutionary stable (Giraud et al., 2002). A further aspect that should be incorporated in future models is that neuronal templates are highly plastic. Ants can habituate to profiles that are entirely different from their own, and tolerate them (Errard et al., 2006; Leonhardt et al., 2007; Florian Menzel, Blüthgen, & Schmitt, 2008). Hence, we need to implement that not only do different cues not necessarily mean perceived chemical differentiation, but also that even perceived chemical differences do no necessarily result in a behavioural response.

References

Tables

Table S1 Effect sizes of t test comparisons between start (tick 99) and end of simulation (tick 1999). The table shows *t* values. Values below zero, which indicate declines over time, are given in red. Each test has df=99 and p < 0.0001.

	Cue diversity	Cue abundance
gestalt	42.1	-92.6
dp	13.9	-59.9
ua	47.2	-121.1
control	-14.8	30.8

Table S2 Effect sizes of t test comparisons between start (tick 99) and end (tick 1999) of the simulation, for different values for profile cost, selection strength and evolvability. The table shows *t* values. Values below zero indicate a decline over time, and are given in red. Each test has df = 99. P values are always < 0.0001 unless stated otherwise.

		Selection strength			Profile cost			Evolvability		
	parameter	0.1	0.4	0.8	20	40	60	1	5	20
Cue diversity	U-absent	20.9	48.9	92	46.4	50.5	56.1	-25.8	50.9	107.1
	Ov.sim.	7.3	41.5	48.7	28.8	36.1	51.9	-25.2	46.7	100.7
	D-present	-3.4	15.7	-0.45**	-12.7	14.4	41.6	-20.5	16.4	112.4
	Control	-1.9*	-13.4	-23.3	-13.5	-14.4	-13.5	-20.9	-15.8	-3.2
Cue abun-	U-absent	-107.1	-139.1	-151.5	-110.6	-108.2	-119.9	-89.6	-110.7	-61.6
	Ov.sim.	-66.3	-89.5	-114.2	-75.6	-98.6	-107.1	-77.6	-97.3	-51.3
	D-present	-63.6	-60.8	-46	47.8	-56.5	-121.1	-68.9	-54.6	-27.2
	Control	12.5	26.9	19.2	27.7	22.3	28.3	3.1***	26.5	71.3

*p = 0.061; **p = 0.66 ***p = 0.0022

Figure legends

Fig. 1 Cue diversity (upper panels) and cue abundance (lower panels) for the three recognition models and the control experiment. The left-hand graphs show mean and standard error for cue diversity and cue abundance (respectively). The right-hand graphs show the values at the end of the simulations (tick 1999), with mean and standard deviation (not standard error). Plots with same letters do not differ significantly according to model summaries.

Fig. 2 Effects of selection strength. The graphs show mean and standard deviation of cue diversity (upper panels) and cue abundance (lower panels) at the end of the simulations (tick 1999), for different levels of selection strength. The four panels depict values for the recognition models: *U-absent, overall similarity, D-present,* and *control.*

Fig. 3 Effects of profile costs. The graphs show mean and standard deviation of cue diversity (upper panels) and cue abundance (lower panels) at the end of the simulations (tick 1999), for different levels of profile costs. The four panels depict values for the recognition models: *U-absent, overall similarity, D-present,* and *control.*

Fig. 4 Effects of evolvability. The graphs show mean and standard deviation of cue diversity (upper panels) and cue abundance (lower panels) at the end of the simulations (tick 1999), for different levels of evolvability. The four panels depict values for the recognition models: *U-absent, overall similarity, D-present,* and *control.*

Supplementary Figures

Fig. S1 Temporal trajectories (tick 99 to 1999) for different levels of selection strength. The graphs show cue diversity and cue abundance (mean and standard error) for different levels of selection strength.

Fig. S2 Temporal trajectories (tick 99 to 1999) for different levels of profile cost. The graphs show cue diversity and cue abundance (mean and standard error) for different levels of selection strength.

Fig. S3 Temporal trajectories (tick 99 to 1999) for different levels of evolvability. The graphs show cue diversity and cue abundance (mean and standard error) for different levels of selection strength.