Invited Review

Queen pheromones and reproductive division of labor: a meta-analysis

Luke Holman

School of BioSciences, University of Melbourne, Royal Parade, Melbourne, VIC 3010, Australia

Our understanding of chemical communication between social insect queens and workers has advanced rapidly in recent years. Several studies have identified chemicals produced by queens and other fertile females that apparently induce sterility in other colony members. However, other experiments produced nonsignificant results, leading some to argue either that earlier reports were mistaken, or that some queen pheromones only work in specific contexts. Here, I review the experimental evidence using meta-analysis, and show that there is near-universal support for the hypothesis that fertility-related chemicals cause sterility regardless of context; studies finding otherwise can be explained most parsimoniously as false negatives. Additionally, queen pheromone experiments that were not performed blind recorded much stronger effect sizes, suggesting bias. I conclude by highlighting several outstanding questions in the field, and by offering recommendations for future studies.

Key words: **Apis mellifera, blind experiments, effect size, eusociality, fertility signals, primer pheromone.**

INTRODUCTION

Queen pheromones are chemical signals that characterize queens and other reproductive individuals in social insects, and are thought to be crucial to the regulation of reproductive division of labor (e.g., Slessor et al. 2005; Le Conte and Hefetz 2008; Van Oystaeyen et al. 2014; Oi et al. 2015). These signals have a multitude of effects, including attracting workers (Slessor et al. 1988), eliciting submissive (Smith et al. 2016) or aggressive (Smith et al. 2009) responses, inhibiting reproduction in workers (Van Oystaeyen et al. 2014) and other queens (Vargo 1992; Holman et al. 2013a), and altering workers' capacity for learning (Vergoz et al. 2007). Queen pheromones appear to be an honest signal that advertises the presence of a healthy reproductive individual, to which workers adaptively respond by continuing to express a worker-like phenotype, as opposed to a manipulation that reduces worker fitness (Keller and Nonacs 1993; Holman 2012; Oi et al. 2015; Peso et al. 2015).

Current evidence suggests that many—possibly all—social insects possess queen pheromones. Experiments with synthetic pheromones have identified bioactive pheromone components in several species, including corbiculate bees, ants, wasps, and termites (e.g., Butler and Fairey 1963; Smith et al. 2009; Holman et al. 2010; Matsuura et al. 2010; Van Oystaeyen et al. 2014; Oi et al. 2016). It is interesting that these taxa all have queen pheromones, because each of them independently evolved eusociality (Peters et al. 2017). Even more remarkably, queen pheromones of different species are

often identical or chemically similar, even in independently-evolved eusocial lineages (Van Oystaeyen et al. 2014). Multiple ants, wasps, and bumblebees have been shown to use cuticular hydrocarbons (CHCs; a nonvolatile blend of hydrocarbons adhering to the body surface) as queen pheromones, particularly certain alkanes, 3-methylalkanes, and alkenes (Van Oystaeyen et al. 2014). By contrast, the well-studied queen pheromone of honeybees (genus *Apis*) is instead thought to be a blend of other chemicals such as keto acids that is secreted from glands, particularly the mandibular gland (Slessor et al. 2005). The only known termite queen pheromone is also thought to be a 2-component blend produced from a gland (Matsuura et al. 2010).

This paper is motivated by outstanding questions over key terminology, the role of learning and context in the response to queen odors, and whether cuticular hydrocarbons can be queen pheromones (Amsalem et al. 2015; Smith and Liebig 2017). Queen pheromones are sometimes called "fertility signals" (e.g., Liebig et al. 2000; Peeters and Liebig 2009; Smith and Liebig 2017), a phrase which has the advantage of highlighting that these chemicals are always (to my knowledge) characteristic of fertile females, rather than queens *per se*. These terms have been used somewhat interchangeably, and so a recent review attempted to distinguish them by redefining fertility signals as those that require "receiver interpretation," and queen pheromones as those that do not (Smith and Liebig 2017). The review did not define "receiver interpretation," but I infer that it means that subordinates learn the reproductive individual's odors by direct association, leading to a conditioned response when the odor is later encountered alone (e.g., on the nest substrate or queen-laid eggs).

Address correspondence to L. Holman. E-mail: luke.holman@unimelb.edu.au.

"Fertility signals" therefore presumably involve associative learning, possibly involving higher brain centres, and might only elicit a response when presented in the correct context (Smith and Liebig 2017). By contrast, "queen pheromones" were proposed to elicit an innate response, presumably involving simple receptor-ligand interactions outside the higher brain. Smith and Liebig (2017) predicted that social insects with small colonies would use fertility signals and those with large colonies would have evolved queen pheromones, and reviewed evidence for this conclusion. Amsalem et al. (2015) made a stronger claim: that CHCs cannot be queen pheromones because they are "so ubiquitous and variable."

To date, most research syntheses on queen pheromones have been qualitative literature reviews (e.g., Monnin 2006; Peeters and Liebig 2009; Kocher and Grozinger 2011; Oi et al. 2015; Smith and Liebig 2017). For example, Smith and Liebig (2017) used "vote counting" (Cooper and Hedges 1994), i.e., tallying studies that support, or did not support, a hypothesis. This approach has limitations. First, in the absence of a formal search method, reviews may miss key literature in a nonrandom fashion. Second, vote counting ignores effect size and statistical power, and recognizes only the presence/absence of a significant result (which depends on sample size). Third, vote counting is especially sensitive to publication bias, also called the "file drawer" problem, biasing the results whenever nonsignificant studies go unpublished. Fourth, there is no quantitative way to weight studies by the quality of their methods (e.g., whether or not they were conducted blind), or test for differences between taxa or types of pheromones.

These problems are largely solved by using a formal literature search followed by meta-analysis (Koricheva et al. 2013). Meta-analysis allows one to build a scientific consensus by amalgamating data from multiple studies, and pooling them to estimate the overall effect size for the experimental condition or relationship of interest. Each study is weighted by its precision, meaning that studies with a good sample size or less variable data contribute more to the overall effect size estimate. Meta-analysis also facilitates detection of publication bias, and allows one to include moderator variables, which can be used to account for biological or methodological differences between studies when building a consensus.

Here, I conduct a meta-analysis that addresses the research question "In eusocial insects, do chemicals that are characteristic of fertile females, such as queens or reproductive workers, reduce the fecundity of other females?." I focus on fecundity, rather than other traits like behavior, because differences in fecundity are fundamental to the "reproductive division of labor" that defines eusociality (Crespi and Yanega 1995). I also focus exclusively on experimental studies because they provide stronger evidence than correlational results, and because the relevant nonexperimental data are covered well by previous reviews (e.g., Monnin 2006; Peeters and Liebig 2009; Kocher and Grozinger 2011; Van Oystaeyen et al. 2014; Oi et al. 2015; Smith and Liebig 2017).

A second aim of this meta-analysis is to test whether the response to queen-derived chemicals varies depending on context, and if they act synergistically with other chemicals. Smith and Liebig (2017)'s hypothesis predicts that queen-derived chemicals will have different effects in different contexts, and may require learning, at least in species with "small" colonies. The hypothesis also predicts that fertility-signaling cuticular hydrocarbons will have no effect when presented in isolation, rather than against a natural background of familiar odors.

Third, it is presently unclear how many queen pheromones are composed of single or multiple component chemicals.

In honeybees, multiple queen pheromone components apparently interact synergistically to attract workers (Slessor et al. 1988; Keeling et al. 2003), and it is often stated that the same is true for fecundity (e.g., Plettner et al. 1996; Hoover et al. 2003; Katzav-Gozansky 2006). In other social insects, multiple different CHCs correlate with fertility, and researchers have tended to assume that all of them are involved in queen–worker communication (Smith et al. 2016). I therefore compared the effect sizes associated with individual chemicals and multicomponent blends.

METHODS
Data collection

I performed a literature search using Web of Science with the following search term: ("queen pheromone*" OR "fertility signal*") OR ("primer pheromone*" AND ovar*), and checked all 332 hits (on 29/08/17). I also checked the reference list for each paper added to the meta-analysis for other suitable papers. Where possible, I obtained the raw data from an online repository or by contacting the authors, or from the paper itself. If the raw data were not available, I searched for descriptive statistics (e.g., means, standard error and sample size for each treatment group) or model parameters (e.g., F or t statistics), which allow calculation of standardized effect size. The Supplementary Material describes data collection for each paper.

Several studies included chemicals that are unrelated to fertility in their experiments, typically as a control for the stimulus of adding a foreign chemical to the nest. Therefore, I recorded whether each chemical tested was a putative fertility signal, based on whether it has been shown to be more abundant in the chemical profiles of queens relative to workers, or fertile individuals relative to nonreproductive ones.

Experiments that are not conducted blind tend to have larger effect sizes due to various types of bias, and it is possible to quantify and partly mitigate this bias using meta-analysis (Holman et al. 2015). Therefore, I also recorded whether each experiment was performed blind; studies were recorded as not blind unless declared otherwise.

Inclusion criteria

The main aim of the meta-analysis is to take stock of the evidence that female fecundity is affected by exposure to chemicals that are produced in greater amounts, or exclusively, by fertile females. I therefore included any study that exposed eusocial insects to 1) chemicals collected from a fertile individual (e.g., CHCs extracted from a queen using solvent), 2) a fertile individual's dead body (or body part), 3) synthetically-produced versions of chemicals produced by fertile individuals, 4) queen-laid eggs that were shown to be coated with queen-like chemicals, or 5) a live queen, provided that a control was available (e.g., a live queen lacking fertility-related CHCs). I only included studies that measured fecundity, or some proxy for fecundity (usually ovary development, or in one case, juvenile hormone titer).

I did not include studies that experimentally tested whether workers respond behaviorally to queen-like chemicals (e.g., Bhadra et al. 2010; Yagound et al. 2015; Smith et al. 2016), because to my knowledge it is not always clear whether these behaviors relate to differences in fecundity. I also excluded experiments that tested for a difference in fecundity between individuals that were housed either with or without a queen, since there are differences between these treatments (e.g., the behavior of the queen) that confound the measurement of the effect of the queen's pheromones. I also

omitted 2 studies (Akre and Reed 1983; Orlova et al. 2013) that attempted to measure the effect of queen-derived chemicals on fecundity, but which had important confounding effects, as well as 5 studies that did not present enough data for me to calculate effect size (see Supplementary Material). Six out of these 7 omitted studies concluded that fertile queens produce sterilizing chemicals, while the seventh (Willis et al. 1990) found a nonsignificant trend in the same direction.

Calculation of effect size

If I could only obtain summary or model statistics, I used them to calculate standardized effect size (presented as a log odds ratio), via R's *compute.es* package. When the raw data were available, I calculated effect size and its 95% confidence limits using contrasts from an appropriate statistical model; this allowed me to remove some of the variation in effect size stemming from differences in statistical methodology, and to correct a few shortcomings in earlier studies' analyses (see Supplementary Material for details of each specific case). The response variable was usually the number of fertile and sterile workers in each treatment, and so I used a binomial generalized linear model (GLM), or, for studies in which the experiment was performed across multiple colonies, a binomial generalized linear mixed model (GLMM; colony was treated as a random factor). If covariates such as body size or colony size were provided in the original study, I included them in the model when calculating effect size, provided the covariate(s) significantly improved model fit. There was no detectable difference between effect sizes calculated with or without covariates (Effect of moderator: $P = 0.39$).

Several studies scored worker ovary activation on an ordinal scale with 3 or more levels (e.g., 0 = undeveloped, 1 = slightly developed, … 4 = highly developed), and then treated the data as a continuous variable when running a *t*-test or ANOVA. However, this is incorrect because category 4 is not "twice as developed" as category 2. I therefore converted this type of data to a binary scale using my best judgment (see Supplementary Material), and analyzed the data with binomial GLM or GLMM.

If a study recorded multiple measurements of worker fecundity, I picked the measurement that is likely to be the best predictor of progeny production or oviposition rate. For example, if the authors counted the number of eggs laid and also performed dissections to measure ovary development, I used the egg laying data. Occasionally, I broke this rule because the best measure of fecundity had a much lower sample size than the second-best measure (e.g., some studies counted egg production for <10 colonies, but measured ovary activation for hundreds of workers). Some individual papers included multiple independent experiments fitting the inclusion criteria, in which case I included all the experiments.

Meta-analysis

I performed a mixed effects meta-regression using the *rma.mv* function from the *metafor* package for R. Each effect size was weighted by the inverse of its standard error. I included 3 moderator variables: *Taxon*, *Fertility signal*, and *Blindness*. *Taxon* had 6 levels: Ant, Bumblebee, Honeybee, Stingless bee, Termite, and Wasp. *Fertility signal* had 2 levels (yes and no), which described whether the focal chemical is produced in greater amounts by fertile females than infertile ones. *Blindness* also had 2 levels, and describes whether the experiment was declared to have been conducted blind. "Experiment" was included as a random effect. I obtained predicted values (and their 95% confidence limits) for the mean effect

size for each combination of moderator variables using the *predict. rma* function. I calculated I^2, the percentage of residual variation that is due to heterogeneity among the effect sizes as opposed to sampling variance, following Nakagawa and Santos (2012). A high value of I^2 implies that the original studies differed substantially in their true effect sizes.

Is the response to queen pheromones learned or context-dependent?

To examine the evidence that 1) the response to queen pheromones varies based on context, and 2) that the response to queen pheromone is learned, I searched the list of recovered papers for experiments that presented a putative queen pheromone in an experimental design that manipulated context (broadly defined) or the opportunity for learning. Additionally, I tested whether individual pheromone components are effective at reducing fecundity using the meta-analysis; a positive result is inconsistent with the hypothesis that queen pheromones are only effective when presented against the correct chemical background (e.g., a familiar profile of hydrocarbons that signal colony membership; Smith and Liebig 2017).

Sterility-regulating queen pheromones: one component or many?

Honeybee "queen mandibular pheromone" (QMP) is a blend of 5 compounds that is sold commercially for beekeeping purposes, and which is thought to have synergistic effects on worker behavior and possibly fecundity (see Introduction for details). The major component of the blend, 9-ODA (9-oxodec-trans-2-enoic acid), also appears to be bioactive when presented in isolation (e.g., Butler and Fairey 1963). I therefore sought to test whether the effect sizes associated with pheromone blends (such as QMP) are stronger than the effect sizes associated with 9-ODA alone, as predicted if honeybee queen pheromones are additive or synergistic. Additionally, I searched for experiments that presented bees with individual chemicals other than 9-ODA. In nonhoneybee species, I similarly compared the efficiency of individual chemicals and blends to inhibit fecundity, and assessed the evidence that more than one chemical affects fecundity when presented alone.

RESULTS

Overview of the dataset

Table 1 shows the number of publications included and effect sizes recovered, and the complete dataset collected for the meta-analysis is described in Supplementary Table S1. The earliest relevant honeybee experiment was from 1954 (Figure 1). Ants (Formicidae) and bumblebees (*Bombus*) were first studied around 1980, followed by higher termites (Rhinotermitidae) in 2010, paper wasps (*Polistes*) in 2007, and yellowjacket wasps (Vespidae) and stingless bees (Meliponinae) in 2014. So far, 16 different species have been studied experimentally: 7 ants, 2 Vespid wasps, 2 honeybees, 2 bumblebees, one stingless bee, one Polistid wasp, and one termite. Almost all the honeybee studies used *Apis mellifera*, though I found a single experiment on *A. cerana*. The ant, bumblebee and wasp studies were more evenly split between species.

In honeybees and termites, all the experiments used chemicals other than cuticular hydrocarbons, or presented entire queens or queen solvent extracts. With one exception, the ant, wasp, stingless bee and bumblebee experiments focused on cuticular hydrocarbons,

or used whole queen extracts; the exception was Van Oystaeyen et al. (2014), which presented *Bombus terrestris* bumblebees with 4 synthetic queen-typical esters, none of which significantly affected fecundity. Eighty-three of the effect sizes came from measurements of the frequency of ovary activation, while the remaining 34 effect sizes used 8 other response variables (e.g., number of eggs laid, or time until egg laying).

Results of the meta-analysis

Studies varied greatly in their precision, though the great majority of effect sizes suggested that fertility-related chemicals significantly reduced fecundity (Figures 2 and 3). Accordingly, the average effect size was negative for all 6 taxa: significantly so for honeybees, termites, and ants (Table 2; Figure 3). For bumblebees, the results were mixed, and there have been rather few experiments on wasps and stingless bees.

Table 1

The number of effect sizes, publications, publications that used blind methods, and species covered by the meta-analysis

Quantity	*n*
Effect sizes	117
Experiments	55
Publications	44
Blind experiments	17
Blind publications	12
Unique species	16
Effect sizes (ants)	33
Effect sizes (honeybees)	47
Effect sizes (bumblebees)	26
Effect sizes (wasps)	6
Effect sizes (termites)	4
Effect sizes (stingless bees)	1

Note that some publications contain multiple independent experiments.

Effect sizes differed significantly between taxa (Table 2). The strongest effect sizes were found in honeybees and termites, while bumblebees had the weakest effect sizes. There was also a borderline nonsignificant trend for fertility-related chemicals to have larger effects on recipient fecundity, relative to chemicals that are unrelated to fertility (Table 2). Chemicals that were not correlated with fecundity never had a statistically significant effect on the fecundity of recipients in any study (in 7 experiments), while fertility-associated chemicals usually did (Figure 2).

Worryingly, I found that only 17 out of 55 (31%) of experiments were conducted blind, and that nonblind results had a log odds ratio that was stronger (i.e., more negative) by 0.77 (Table 2), which is considered large (Koricheva et al. 2013). The use of blind methods varied between taxa: only 1 out of 25 honeybee experiments, and none of the termite experiments, were performed blind, while most of the experiments on ants, wasps and bumblebees were blind (Supplementary Table S2).

Because the frequency of blind studies differs between taxa (i.e., there is some collinearity among the moderator variables), my model might mismeasure the magnitude of the differences between taxa, or the effect of not working blind. This has implications for the biological interpretation of the meta-analysis results. For example, the glandular pheromones of advanced eusocial species like honeybees are regarded as having stronger effects on worker fecundity than the CHC-based pheromones of ants and wasps by some (e.g., Amsalem et al. 2015), but it is hard to compare effect sizes across taxa because the honeybee and termite experiments may have incurred more observer bias.

A substantial percentage of the residual variance was explained by heterogeneity in effect size rather than sampling variance ($I^2 = 70\%$), suggesting that true effect sizes varied considerably among experiments. Inspection of the data revealed that honeybee effect sizes calculated from treatment means (as opposed to those calculated from the raw data, or from F, t, or X^2 statistics) were
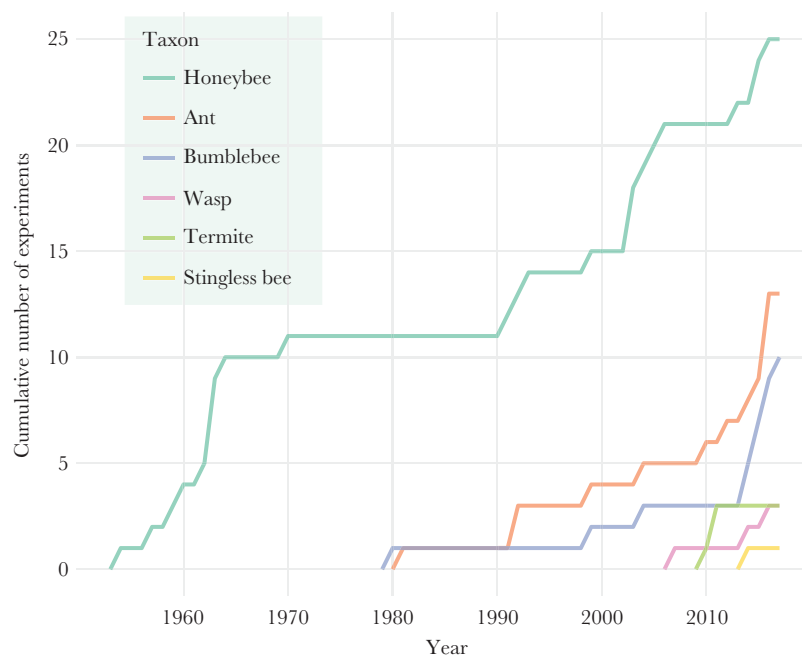

Figure 1

Cumulative number of experiments examining the effect of queen chemicals on fecundity per year in each taxon (note that some publications contained more than one experiment).
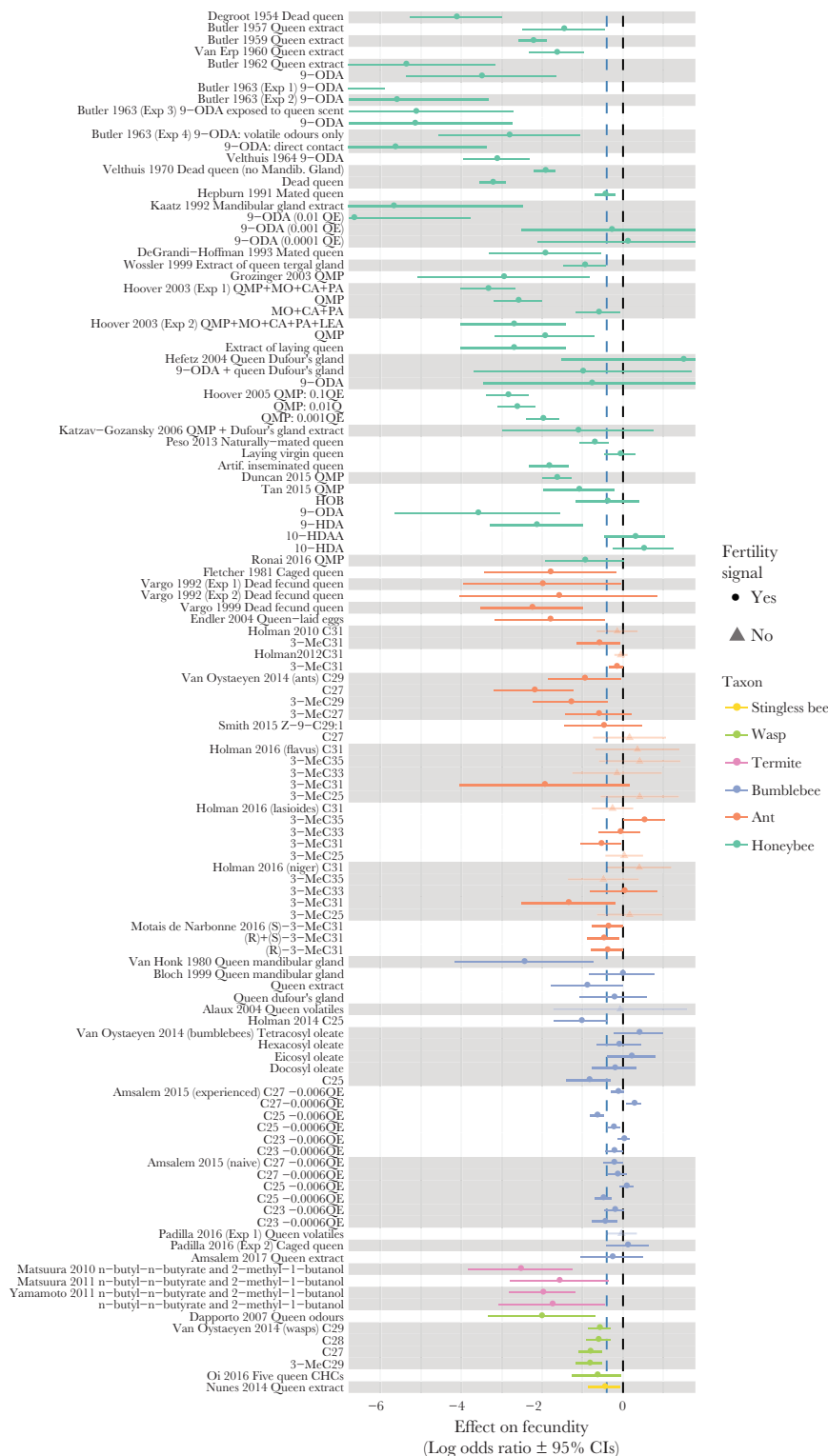
Figure 2

All 117 effect sizes used in the meta-analysis, showing that most fertility-related chemicals tested so far reduce the fecundity of recipients (denoted by effect size <0). The grey and white shading identifies effect sizes that come from the same experiment, and the y-axis gives the first author's name and the publication date of the study, as well as the chemical being tested. Each effect size is calculated relative to a control, typically a solvent-only treatment. The black dashed line marks an effect size of zero, and the blue dashed line marks the point at which effect sizes are conventionally considered large (equivalent to Cohen's $d = 0.5$). Semitransparent triangles mark effect sizes relating to chemicals that do not correlate with fecundity, which are sometimes used as controls in experiments that also test a fertility signal. See the Supplementary Material for complete references, a description of each study, and complete information on how every effect size was obtained. QE stands for "queen equivalents", i.e., the pheromone dose relative to the average quantity possessed by a single queen.
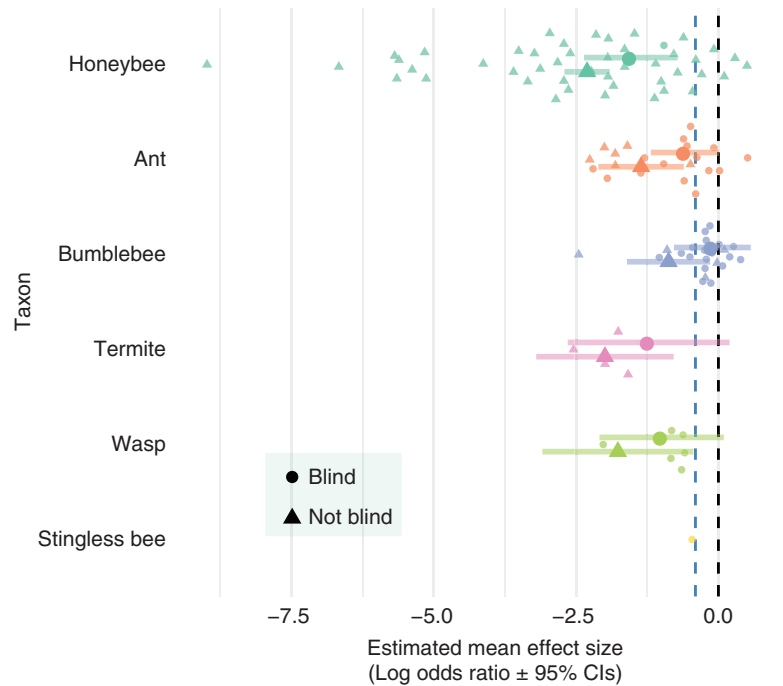
Figure 3

The points with error bars show the estimated mean effect size for each taxon, among blind and nonblind studies, for fertility-signaling chemicals (i.e., the predicted values derived from the meta-analysis). The points lacking error bars are the individual effect sizes, also shown in Figure 2. Note that the apparent lack of an effect for bumblebees is driven by one large experiment finding that queen-derived esters have no effect on fecundity; when the esters are excluded, effect size becomes significantly negative (see main text).

Table 2

Effects of moderator variables in the meta-analysis

Parameter	Effect (log odds ratio)	SE	z	p
Intercept	-1.11 (-2.04 to -0.17)	0.48	-2.31	0.021
Taxon: Ant	0.95 (0.12 to 1.78)	0.42	2.25	0.025
Bumblebee	1.43 (0.62 to 2.25)	0.42	3.45	0.0006
Stingless bee	1.08 (-0.92 to 3.07)	1.02	1.06	0.290
Termite	0.31 (-0.96 to 1.58)	0.65	0.48	0.628
Wasp	0.54 (-0.82 to 1.91)	0.70	0.78	0.437
Fertility signal: Yes	-0.43 (-0.87 to 0.007)	0.22	-1.93	0.054
Blind: No	-0.77 (-1.52 to -0.015)	0.38	-2.00	0.046

There are 3 moderators in the model—Taxon, Fertility signal, and Blind—which have 6, 2, and 2 levels, respectively. The intercept shows the estimated overall effect size at the reference levels of each moderator, namely Taxon: honeybee, Fertility signal: No, and Blind: Yes. The other effect sizes indicate the effect of changing one of these moderators; for example, effect sizes associated with fertility signals or nonblind experiments tend to be more negative. Numbers in parentheses are 95% confidence limits.

far more heterogeneous than effect sizes from any other source (Supplementary Figure S1; effect size range: −9.0 to 1.5). As a sensitivity analysis, I reran the meta-analysis with these data removed and obtained very similar results, as well as considerably lower heterogeneity (Supplementary Material; I^2 = 43%). Qualitatively, the only difference in the results was that the moderator "Fertility signal" became statistically significant (P = 0.014).

Lastly, I noticed that all of the bumblebee treatments that did not include cuticular hydrocarbons (i.e., queen volatile odors, or esters such as hexacosyl oleate) appeared to have below-average effect sizes (Figure 2). As a *post-hoc* exploratory analysis, I ran a meta-analysis on just the bumblebee data, after eliminating all effect sizes that did not include cuticular hydrocarbons. This meta-analysis included

20 effect sizes from 8 experiments, all of which presented either synthetic fertility-related hydrocarbons, queens, or queen cuticular washes. The mean effect size was −0.24 (95% CIs = −0.40 to −0.080, P = 0.0035, I^2 = 0.00%). Thus, fertility-associated hydrocarbons appear to reduce fecundity in bumblebees.

Relationship between sample size and statistical power

The large error bars in Figure 2 highlight that many studies were not able to precisely measure effect size. For example, the nonsignificant result from Smith et al. (2015), used by Smith and Liebig (2017) as evidence that fertility-associated CHCs have no effect on fecundity when presented in the wrong context, had effect size 95% confidence limits that ran from −1.45 to 0.48. This means that the study's sample size (n = 13) was not high enough to determine whether queen-like CHCs had no effect, a strong negative effect, or even a strong positive effect on worker fecundity.

To illustrate that inadequate sample size, as opposed to biologically relevant factors, could be the main reason why some queen pheromone experiments have produced nonsignificant results, I compared the sample size of the significant and nonsignificant effects (Figure 4). All but one of the studies that found no significant effect of fertility-associated cuticular hydrocarbons on fecundity had a sample size of n = 8–45 per treatment, i.e., well below the average sample size. The nonsignificant data points with n > 100 come from Holman et al. (2016), which produced a mixture of significant and nonsignificant results (for CHCs with typical or atypical chain length, respectively).

The results of this meta-analysis can be used to help plan future experiments. One can readily calculate the sample size required to obtain a specific probability (power) of detecting an effect of the
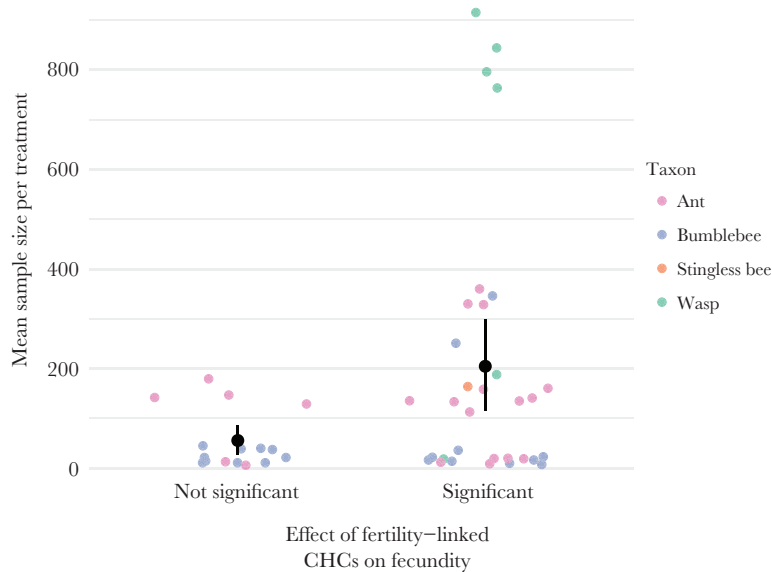
Figure 4

Average sample size per treatment (±95% confidence limits) for every effect size describing the effect of fertility-related cuticular hydrocarbons (CHCs) on fecundity, for significant and nonsignificant studies. Most of the nonsignificant effects had a lower than average sample size ($n = 8–45$), and all were well below the level needed to achieve good statistical power given the effect sizes in Figure 3 (see Supplementary Table S4). This suggests that qualitative differences between studies' conclusions can most parsimoniously be explained by false negatives in underpowered experiments, rather than differences in biology. The 4 nonsignificant results with $n > 100$ are from Holman et al. (2016), which presented ants with CHCs of atypical chain length.

same size as those estimated for queen pheromones in each taxon (Figure 3). For example, in order to have 80% power to reject the null hypothesis with $\alpha = 0.95$ when the true effect size is large as the average for ant fertility signals (i.e., Log odds ratio $= −0.59$), one needs to have 145 samples in each treatment group. Supplementary Table S3 summarizes the sample sizes needed to reliably detect effect sizes as large as the taxon-specific averages shown in Figure 3.

No direct evidence for context-dependent responses or learning

I did not find any experiments that provided an unconfounded test of the hypothesis that queen pheromones have context-specific effects on fecundity, or that the response to pheromone must be learned. The most relevant study was Amsalem et al. (2015), who presented queen-like CHCs to workers that had previously encountered a queen, and workers that had not done so. However, the naive and experienced workers differed in several other respects, as explained by Holman et al. (2017), confounding measurement of the effect of learning. Moreover, the naïve workers also showed a statistically significant reduction in fecundity in response to queen pheromone (Figure 2). Additionally, Smith et al. (2015) recorded the frequency of submissive behavior in workers exposed to a synthetic queen pheromone presented against a background of same-population or different-population worker hydrocarbons. However, submissive behavior might not translate into differences in fecundity, and the experiment was not performed blind.

Additionally, there was abundant evidence (Figures 2 and 3) that individual queen-like chemicals can reduce fecundity even when presented in isolation in unnatural conditions, which does not support the hypothesis that these pheromones require a specific context or chemical background to function.

Sterility-regulating queen pheromones: one component or many?

Four studies presented honeybees with different combinations of pheromones, allowing them to test experimentally for synergy (summarized in Supplementary Figure S2). Butler et al. (1962) exposed workers to solvent-extracted queen chemicals, 9-ODA alone, or a control, and found that the queen extract had the strongest effect. However, it is unclear whether the difference was statistically significant, and it seems likely that the concentration of 9-ODA differed between the 2 treatments, confounding the experiment. Second, 2 studies compared 9-ODA to the 5-compound blend termed queen mandibular pheromone (QMP), and found no difference in their effects on worker fecundity (Kaatz et al. 1992; Tan et al. 2015). Lastly, Hefetz and Katzav-Gozansky (2004) tested whether QMP had a stronger effect when presented alongside secretions from the queen's Dufour's gland, and found no significant difference, though the sample size was $n = 5$. Finally, 2 additional papers measured fecundity in bees presented with 9-ODA alone (Butler and Fairey 1963; Velthuis and van Es 1964), and recorded effect sizes that were similar to or higher than those reported for QMP or whole-queen extracts (Supplementary Figure S2).

The Tan et al. (2015) study also found that workers exposed to another component of QMP, 9-HDA (9-hydroxy-(E)2-decenoic acid), displayed significantly reduced fecundity. To my knowledge, this is the only study to show that a specific compound other than 9-ODA significantly affects fecundity, and thus provides the most direct evidence that honeybee queens indeed produce more than one chemical affecting worker sterility. Additionally, Wössler and Crewe (1999) concluded that solvent extracts from queens' ter-gal glands induced sterility in groups of caged workers. Although it seems possible that the tergal gland extracts contained traces of

chemicals from the mandibular gland (e.g., from self-grooming), this experiment suggests that tergal gland secretions (which are fertility-signaling hydrocarbons; Smith et al. 1993) might also be pheromones that affect worker sterility.

In nonhoneybee species, 6 experiments have tested whether multiple different fertility-related CHCs inhibited fecundity (Figure 2). In *Vespula vulgaris* wasps, *Cataglyphis iberica* ants, and *Bombus impatiens* bees, more than one chemical significantly affected fecundity (Van Oystaeyen et al. 2014; Amsalem et al. 2015). In 3 species of *Lasius* ants (Holman et al. 2016), multiple CHCs were tested but only one had a significant effect. No experiments have yet compared the effect of a blend of chemicals to that of the individual components, so there is presently no evidence for additive or synergistic interactions among pheromone components.

Publication bias appears weak

Figure 5 shows a funnel plot of the effect sizes in the meta-analysis. There was a slight over-representation of low-powered studies with stronger-than-average negative effects on worker fecundity, implying that very small studies need to find significant results in order to be published. Because the effect sizes in Figures 2 and 3 are mostly very large, the overall conclusions of the meta-analysis are unlikely to be driven by publication bias.

DISCUSSION

The meta-analysis revealed that over the last 60 years, at least 44 publications have experimentally tested the hypothesis that chemicals associated with fertile social insect females (typically queens) affect the fecundity of other females (typically workers). In ants, wasps, one species of higher termite and especially honeybees, the evidence appears definitive: multiple independent results support the existence of queen pheromones that modulate fecundity. In bumblebees, the evidence was weaker, though it is notable that the 3 largest studies (Holman 2014; Van Oystaeyen et al. 2014; Amsalem et al. 2015), which were conducted blind by 3 different research teams, all found evidence that fertility-signaling CHCs significantly reduce worker fecundity.

The meta-analysis suggested that the glandular pheromones of honeybees and termites might have stronger effects on workers than the CHC-based pheromones of ants, wasps and other bees. However, it is difficult to ascertain the size of the difference or even to be confident that it is genuine, because almost all the honeybee and termite research was not conducted blind, unlike studies of the other species. Worryingly, nonblind studies reported substantially stronger effects than did blind studies (as found many times previously; reviewed in Holman et al. 2015). This result suggests that nonblind studies incurred various types of bias (e.g., observer bias) that exaggerated the effect of queen pheromones.

When testing the theory that fertility-related CHCs only affect worker fecundity when presented in the correct context (Smith and Liebig 2017), I first searched for studies that presented CHCs in multiple contexts and then tested for context-dependent effects on fecundity. I didn't find any such studies, and so it appears that the theory presently lacks direct evidence. Second, I tested the theory's prediction that fertility-related CHCs should have no effect on fecundity when presented in isolation. This prediction was not supported: the great majority of experiments that presented individual
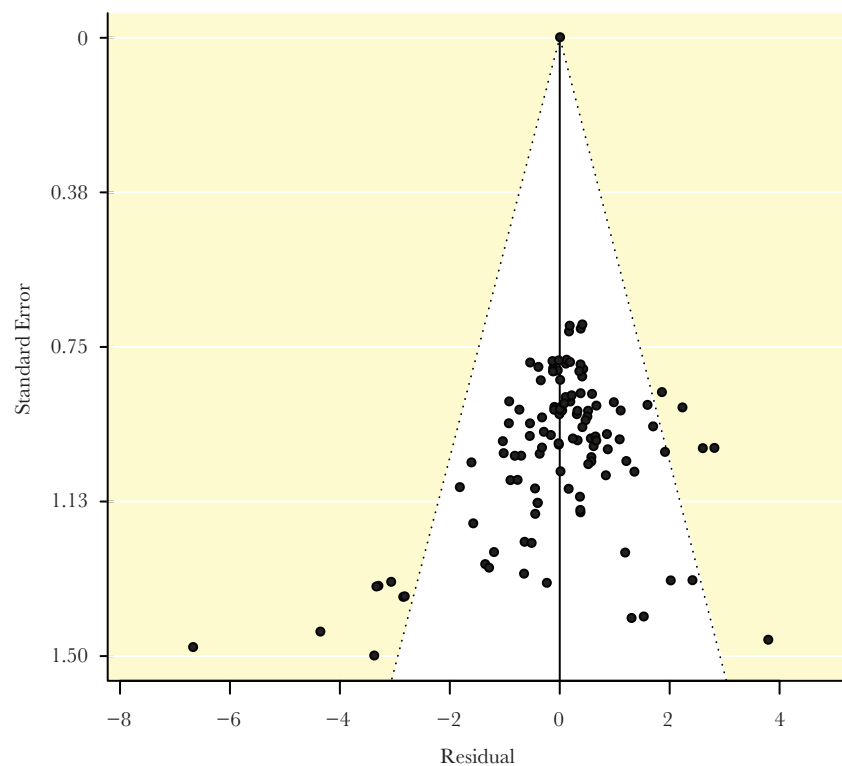

Figure 5

Funnel plot, illustrating that publication bias is weak or absent. There is a slight shortage of low-powered studies finding a more positive (i.e., weaker inhibitory) effect of queen pheromones on fecundity, consist with publication bias against low-powered, nonsignificant studies. The plot shows the residual for each effect size from the meta-analysis plotted against its standard error, and the yellow region denotes samples that fall outside the 95% confidence intervals expected based on the overall distribution of effects (plot produced using the *metafor* R package).

fertility-related CHCs recorded strong, statistically significant effects on fecundity, and the few that did not tended to have low power. For example, the null result cited by Smith and Liebig as evidence for their theory had a sample size of 13, which I estimate to be 10-fold less than the sample size required to reach 80% power in an experiment on ant queen pheromones (Supplementary Table S3). Thus, it seems more parsimonious to regard the few published non-significant results as false negatives, rather than evidence that particular taxa exhibit complex, context-dependent responses.

I also tested whether queens produce a blend of pheromones that act together synergistically to inhibit worker fecundity, as is often claimed. So far, 4 honeybee studies have experimentally tested for synergy, and 0 out of 4 found clear evidence for it; additionally, 5 out of 5 studies testing the main component of queen mandibular pheromone, 9-ODA, found that it has a strong effect on worker fecundity when presented alone. However, pheromones do appear to have synergistic effects on honeybee "retinue" behavior (Slessor et al. 1988; Keeling et al. 2003), so it is plausible that synergistic effects on fecundity exist but have thus far escaped detection. Additionally, one study found that another component of the queen mandibular pheromone, 9-HDA, affected worker fecundity in *A. cerana* (Tan et al. 2015), and another concluded that extracts of queens' tergal glands (Smith et al. 1993) caused sterility in *A. mellifera* workers (Wossler and Crewe 1999). 9-HDA is a direct biosynthetic precursor of 9-ODA (Plettner et al. 1996), and is produced in higher quantities by fertile queens (Strauss et al. 2008). The effect on worker fertility of 9-HDA, or any individual chemical other than 9-ODA, has to my knowledge never been tested in *A. mellifera*. The finding that queen tergal glands inhibit fecundity raises the interesting possibility that honeybees possess CHC-based queen pheromones, much like all other social Hymenoptera studied so far, in addition to glandular queen pheromones. This hypothesis remains to tested directly, though tellingly, the hydrocarbon profile of fertile queen honeybees is distinguished from that of workers by an excess of linear alkanes (Babis et al. 2014), which have been experimentally identified as queen pheromones in ants, bumblebees, and wasps.

The idea that the honeybee pheromone is a complex blend has been used to argue that it is a "manipulative" adaptation, with which queens sterilize workers against the workers' fitness interests (Katzav-Gozansky 2006). The multicomponent nature of the pheromone is hypothesized to be the result of a chemical arms race, in which queens evolved new pheromone components to restore control each time that workers evolved resistance to sterilization. Settling whether the pheromone really is a blend could therefore shed light on queen pheromone evolution.

Suggestions for future work

In addition to the knowledge gaps described above, many eusocial insect clades presently have no experimentally-identified queen pheromones that affect fecundity, including lower termites, Stenogastrine wasps, Allodapine and Halicitid bees, and most ant subfamilies (including large clades like the Ponerinae and Dolichoderinae). In *Polistes* wasps (Dapporto et al. 2007) and stingless bees (Nunes et al. 2014), queen-derived chemical mixtures were found to affect worker fecundity, but the compounds involved have yet to be identified. It will be interesting to determine how the chemicals used in queen-worker communication vary across the phylogeny, and to determine the ecological and evolutionary forces that determine the identity and number of chemicals that make up the pheromone. Existing data suggest that queen pheromones are

slow-evolving (Brunner et al. 2011; Holman et al. 2013b), but it is clear that not all taxa use the same pheromones. For example, the CHC 3-MeC$_{31}$ is a queen pheromone throughout the ant genus *Lasius*, but this chemical is absent in many other eusocial ants and bees (Holman et al. 2013b). To maximize the power of phylogenetic analyses, researchers could focus on experimentally identifying queen pheromones in pairs of sister taxa which differ in some trait of interest, e.g., social complexity, queen-worker differentiation, queen number, or the prevalence of inquiline social parasites that mimic the queen, and then test for correlated evolution.

It is also unclear how queen pheromones achieve their effects. Recent work (Slone et al. 2017; Pask et al. 2017) reaffirmed earlier results (e.g., d'Ettorre et al. 2004; Ozaki et al. 2005; Holman et al. 2010) that ants perceive cuticular hydrocarbons via their antennae, and identified odorant binding proteins that bind hydrocarbons. An odorant binding protein that binds the major honeybee queen pheromone component, 9-ODA, has also been discovered (Wanner et al. 2007). However, to my knowledge it is unclear what happens after the pheromone binds to its receptor. A pioneering study by Grozinger et al. (2003) used microarrays to identify many genes whose expression changes in response to queen pheromone in the honeybee, and future studies (either in social insects or in models such as *Drosophila;* Camiletti et al. 2014) will hopefully elucidate the genetic and physiological cascade that begins with the pheromone receptor and ends with phenotypic change. Determining the mechanisms by which queen pheromones act will help shed light on how they originated, evolved, and diversified.

The ant *Harpagnethos saltator* has become a model for mechanistic studies of queen pheromone perception (Slone et al. 2017; Pask et al. 2017). Oddly, no published study has attempted to experimentally identify any queen pheromones in this species: the most pertinent study simply documents differences in the CHC profiles of *H. saltator* queens and workers (Liebig et al. 2000). However, Figure 2 highlights that we cannot assume that every chemical produced by queens is a pheromone: several studies have presented multiple candidate queen pheromones, and found that only some of them elicit a detectable response (e.g., Van Oystaeyen et al. 2014). I therefore suggest experimental identification of the active component(s) of the queen pheromone, to maximize the utility of this promising model.

Regarding experimental design, I recommend that experimenters work blind while running experiments and collecting data. Doing so is usually as simple as getting a colleague to relabel each pheromone treatment with a code, which is decoded once all the data are collected. Many queen pheromone studies focus on a fairly subjective response variable, e.g., whether an ovary is developed or undeveloped, leaving room for observer bias when recording data. Working blind also ensures that one does not inadvertently handle the treatment and control groups differently. Another key tenet of experimental design is that one should start with a common pool of individuals, and then randomly allocate them to each treatment (e.g., the control and pheromone-treated groups). One study was excluded from the meta-analysis because the treatment and control were performed in sequence at different times of the year (Orlova et al. 2013), and in another study (Amsalem et al. 2015), workers from the "experienced" and "naive"' treatments also differed in age, size, and colony origin, confounding measurement of the effects of experience (Holman et al. 2017).

Statistical power is also worth careful consideration when designing experiments. Most studies base their conclusions solely on the *P*-value, but statistical significance relies on sample size in addition

to the true effect size (Nakagawa and Cuthill 2007). When sample size is low, a nonsignificant result is uninformative because it has a high probability of being a false negative. This means that it is only worth adding additional treatments and conditions if one can maintain adequate replication. For example, an ambitious study by Amsalem et al. (2015) examined 14 different treatments, but sample size dropped as low as $n = 6$ per treatment. Supplementary Table S4 illustrates that one should aim to have higher replication than this—generally $n > 100$—to have adequate confidence in negative results.

Calculating effect size and its confidence intervals helps the reader to interpret both nonsignificant and significant results (Nakagawa and Cuthill 2007). Nonsignificant results for which the effect size confidence intervals tightly bound zero are less likely to be false negatives than nonsignificant results with very wide confidence intervals, and the confidence intervals associated with a significant result illustrate whether the effect is small or large. Although some queen pheromone studies have used "post-hoc power analysis" for a similar purpose (Amsalem et al. 2015), this method is misleading and should not be used (Levine and Ensom 2001). Lastly, one should regard each study as a contribution to a larger body of evidence rather than a decisive answer, and be mindful that individual experiments have a high likelihood of being false (Ioannidis 2014) when developing hypotheses to explain differences among studies.

CONCLUSIONS

There is very strong evidence that queens and other fertile females produce chemicals that inhibit reproduction in other females, in diverse Hymenoptera and one termite. Although a handful of experiments concluded that ant and bumblebee CHCs are not queen pheromones, the most parsimonious explanation is that these results represent false negatives, rather than cases in which a species has lost or modified its response to queen odors. A simple model whereby queen pheromones bind to pheromone receptors and affect an innate, "hard-wired" response cannot presently be falsified. Alternative hypotheses—e.g., involving learning, or synergistic interactions among multiple olfactory cues—cannot be ruled out either, but they presently lack the experimental support needed to supplant the more parsimonious first hypothesis. The evolution and mechanistic basis of queen-worker chemical communication remains incompletely understood, and I look forward to new developments in this exciting field.

Editor-in-Chief: Leigh Simmons

REFERENCES

Akre RD, Reed HC. 1983. Evidence for a queen pheromone in *Vespula* (Hymenoptera: Vespidae). Can Entomol. 115:371–377.

Amsalem E, Orlova M, Grozinger CM. 2015. A conserved class of queen pheromones? Re-evaluating the evidence in bumblebees (*Bombus impatiens*). Proc Roy Soc B. 282:20151800.

Babis M, Holman L, Fenske R, Thomas ML, Baer B. 2014. Cuticular lipids correlate with age and insemination status in queen honeybees. Insect Soc. 61:337–345.

Bhadra A, Mitra A, Deshpande SA, Chandrasekhar K, Naik DG, Hefetz A, Gadagkar R. 2010. Regulation of reproduction in the primitively eusocial wasp *Ropalidia marginata*: on the trail of the queen pheromone. J Chem Ecol. 36:424–431.

Brunner E, Kroiss J, Trindl A, Heinze J. 2011. Queen pheromones in *Temnothorax* ants: control or honest signal? BMC Evol Biol. 11:55.

Butler CG, Callow RK, Johnston NC. 1962. The isolation and synthesis of queen substance, 9-oxodec-trans-2-enoic acid, a honeybee pheromone. Proc Roy Soc B. 155:417–432.

Butler CG, Fairey EM. 1963. The role of the queen in preventing oogenesis in worker honeybees. J Apic Res. 2:14–18.

Camiletti AL, Awde DN, Thompson GJ. 2014. How flies respond to honey bee pheromone: the role of the *foraging* gene on reproductive response to queen mandibular pheromone. Naturwissenschaften. 101:25–31.

Cooper H, Hedges LV. 1994. Vote counting procedures in meta-analysis. In: The handbook of research synthesis. New York: Russell Sage Foundation. p. 193–214.

Crespi BJ, Yanega D. 1995. The definition of eusociality. Behav Ecol. 6:109–115.

D'Ettorre P, Heinze J, Schulz C, Francke W, Ayasse M. 2004. Does she smell like a queen? Chemoreception of a cuticular hydrocarbon signal in the ant *Pachycondyla inversa*. J Exp Biol. 207:1085–1091.

Dapporto L, Santini A, Dani FR, Turillazzi S. 2007. Workers of a *Polistes* paper wasp detect the presence of their queen by chemical cues. Chem Senses. 32:795–802.

Grozinger CM, Sharabash NM, Whitfield CW, Robinson GE. 2003. Pheromone-mediated gene expression in the honey bee brain. PNAS. 100:14519–14525.

Hefetz A, Katzav-Gozansky T. 2004. Are multiple honeybee queen pheromones indicators for a queen-workers arms race? Apiacta. 39:44–52.

Holman L. 2012. Costs and constraints conspire to produce honest signaling: insights from an ant queen pheromone. Evolution. 66:2094–2105.

Holman L. 2014. Bumblebee size polymorphism and worker response to queen pheromone. PeerJ. 2:e604.

Holman L, Hanley B, Millar JG. 2016. Highly specific responses to queen pheromone in three Lasius ant species. Behav Ecol Sociobiol. 70:387–392.

Holman L, Head ML, Lanfear R, Jennions MD. 2015. Evidence of experimental bias in the life sciences: why we need blind data recording. PLoS Biol. 13:e1002190.

Holman L, Jørgensen CG, Nielsen J, d'Ettorre P. 2010. Identification of an ant queen pheromone regulating worker sterility. Proc Roy Soc B. 277:3793–3800.

Holman L, Lanfear R, d'Ettorre P. 2013b. The evolution of queen pheromones in the ant genus *Lasius*. J Evol Biol. 17:1549–1558.

Holman L, Leroy C, Jørgensen C, Nielsen J, d'Ettorre P. 2013a. Are queen ants inhibited by their own pheromone? Regulation of productivity via negative feedback. Behav Ecol. 24:380–385.

Holman L, van Zweden JS, Oliveira RC, van Oystaeyen A, Wenseleers T. 2017. Conserved queen pheromones in bumblebees: a reply to Amsalem *et al*. PeerJ. 5:e3332.

Hoover SE, Keeling CI, Winston ML, Slessor KN. 2003. The effect of queen pheromones on worker honey bee ovary development. Naturwissenschaften. 90:477–480.

Ioannidis JP. 2014. How to make more published research true. PLoS Med. 11:e1001747.

Kaatz H-H, Hildebrandt H, Engels W. 1992. Primer effect of queen pheromone on juvenile hormone biosynthesis in adult worker honey bees. J Comp Physiol B. 162:588–592.

Katzav-Gozansky T. 2006. The evolution of honeybee multiple queen pheromones: a consequence of a queen-worker arms race? Braz J Morphol Sci. 23:287–294.

Keeling CI, Slessor KN, Higo HA, Winston ML. 2003. New components of the honey bee (*Apis mellifera* L.) queen retinue pheromone. Proc Natl Acad Sci USA. 100:4486–4491.

Keller L, Nonacs P. 1993. The role of queen pheromones in social insects: queen control or queen signal? Anim Behav. 45:787–794.

Kocher SD, Grozinger CM. 2011. Cooperation, conflict, and the evolution of queen pheromones. J Chem Ecol. 37:1263–1275.

Koricheva J, Gurevitch J, Mengersen K. 2013. Handbook of meta-analysis in ecology and evolution. Princeton: Princeton University Press.

Le Conte Y, Hefetz A. 2008. Primer pheromones in social hymenoptera. Annu Rev Entomol. 53:523–542.

Levine M, Ensom MH. 2001. Post hoc power analysis: an idea whose time has passed? Pharmacotherapy. 21:405–409.

Liebig J, Peeters C, Oldham NJ, Markstadter C, Hölldobler B. 2000. Are variations in cuticular hydrocarbons of queens and workers a reliable signal of fertility in the ant *Harpegnathos saltator*? PNAS. 97:4124–4131.

Matsuura K, Himuro C, Yokoi T, Yamamoto Y, Vargo EL, Keller L. 2010. Identification of a pheromone regulating caste differentiation in termites. PNAS. 107:12963–12968.

Monnin T. 2006. Chemical recognition of reproductive status in social insects. Annales Zoologici Fennici. 43:531–549.

Nakagawa S, Cuthill IC. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev. 82:591–605.

Nakagawa S, Santos ESA. 2012. Methodological issues and advances in biological meta-analysis. Evol Ecol. 26:1253–1274.

Nunes TM, Mateus S, Favaris AP, Amaral MFZJ, von Zuben LG, Clososki GC, Bento JMS, Oldroyd BP, Silva R, Zucchi R, et al. 2014. Queen signals in a stingless bee: suppression of worker ovary activation and spatial distribution of active compounds. Sci Rep. 4:7449.

Oi CA, Millar JG, van Zweden JS, Wenseleers T. 2016. Conservation of queen pheromones across two species of Vespine wasps. J Chem Ecol. 42:1175–1180.

Oi CA, van Zweden JS, Oliveira RC, Van Oystaeyen A, Nascimento FS, Wenseleers T. 2015. The origin and evolution of social insect queen pheromones: novel hypotheses and outstanding problems. Bioessays. 37:808–821.

Orlova M, Malka O, Hefetz A. 2013. Virgin honeybee queens fail to suppress worker fertility but not fertility signalling. J Insect Physiol. 59:311–317.

Ozaki M, Wada-Katsumata A, Fujikawa K, Iwasaki M, Yokohari F, Satoji Y, Nisimura T, Yamaoka R. 2005. Ant nestmate and non-nestmate discrimination by a chemosensory sensillum. Science. 309:311–314.

Pask GM, Slone JD, Millar JG, Das P, Moreira JA, Zhou X, Bello J, Berger SL, Bonasio R, Desplan C, et al. 2017. Specialized odorant receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. Nature Comm. 8:297.

Peeters C, Liebig J. 2009. Fertility signaling as a general mechanism of regulating reproductive division of labor in ants. In: Gadau J, Fewell J, editors. Organization of insect societies: from genome to socio-complexity. Cambridge (MA): Harvard University Press.

Peso M, Elgar MA, Barron AB. 2015. Pheromonal control: reconciling physiological mechanism with signalling theory. Biol Rev Camb Philos Soc. 90:542–559.

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, et al. 2017. Evolutionary history of the Hymenoptera. Curr Biol. 27:1013–1018.

Plettner E, Slessor KN, Winston ML, Oliver JE. 1996. Caste-selective pheromone biosynthesis in honeybees. Science. 271:1851–1853.

Slessor KN, Kaminski L-A, King GGS, Borden JH, Winston ML. 1988. Semiochemical basis of the retinue response to queen honey bees. Nature. 332:354–356.

Slessor KN, Winston ML, Le Conte Y. 2005. Pheromone communication in the honeybee (*Apis mellifera* L.). J Chem Ecol. 31:2731–2745.

Slone JD, Pask GM, Ferguson ST, Millar JG, Berger SL, Reinberg D, Liebig J, Ray A, Zwiebel LJ. 2017. Functional characterization of odorant receptors in the ponerine ant, *Harpegnathos saltator*. PNAS. 114:8586–8591.

Smith AA, Hölldober B, Liebig J. 2009. Cuticular hydrocarbons reliably identify cheaters and allow enforcement of altruism in a social insect. Curr Biol. 19:78–81.

Smith AA, Liebig J. 2017. The evolution of cuticular fertility signals in eusocial insects. Curr Opin Insect Sci. 22:79–84.

Smith AA, Millar JG, Suarez AV. 2015. A social insect fertility signal is dependent on chemical context. Biol Lett. 11:20140947.

Smith AA, Millar JG, Suarez AV. 2016. Comparative analysis of fertility signals and sex-specific cuticular chemical profiles of *Odontomachus* trap-jaw ants. J Exp Biol. 219:419–430.

Smith RK, Spivak M, Taylor OR Jr, Bennett C, Smith ML. 1993. Maturation of tergal gland alkene profiles in European honey bee queens, *Apis mellifera* L. J Chem Ecol. 19:133–142.

Strauss K, Scharpenberg H, Crewe RM, Glahn F, Foth H, Moritz RFA. 2008. The role of the queen mandibular gland pheromone in honeybees (*Apis mellifera*): honest signal or suppressive agent? Behav Ecol Sociobiol. 62:1523–1531.

Tan K, Liu X, Dong S, Wang C, Oldroyd BP. 2015. Pheromones affecting ovary activation and ovariole loss in the Asian honey bee *Apis cerana*. J Insect Physiol. 74:25–29.

Van Oystaeyen A, Oliveira RC, Holman L, van Zweden JS, Romero C, Oi CA, d'Ettorre P, Khalesi M, Billen J, Wäckers F, et al. 2014. Conserved class of queen pheromones stops social insect workers from reproducing. Science. 343:287–290.

Vargo EL. 1992. Mutual pheromonal inhibition among queens in polygyne colonies of the fire ant *Solenopsis invicta*. Behav Ecol Sociobiol. 31:205–210.

Velthuis HJ, van Es J. 1964. Some functional aspects of the mandibular glands of the queen honeybee. J Apic Res. 3:11–16.

Vergoz V, Schreurs HA, Mercer AR. 2007. Queen pheromone blocks aversive learning in young worker bees. Science. 317:384–386.

Wanner KW, Nichols AS, Walden KKO, Brockmann A, Luetje CW, Robertson HM. 2007. A honey bee odorant receptor for the queen substance 9-oxo-2-decenoic acid. PNAS. 104:14383–14388.

Willis LG, Winston ML, Slessor KN. 1990. Queen honey bee mandibular pheromone does not affect worker ovary development. The Canadian Entomologist. 122:1093–1099.

Wossler TC, Crewe RM. 1999. Honeybee queen tergal gland secretion affects ovarian development in caged workers. Apidologie. 30:311–320.

Yagound B, Gouttefarde R, Leroy C, Belibel R, Barbaud C, Fresneau D, Chameron S, Poteaux C, Châline N. 2015. Fertility signaling and partitioning of reproduction in the ant *Neoponera apicalis*. J Chem Ecol. 41:557–566.