PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Zareyan S, Otto SP, Hauert C. 2019 A sheep in wolf's clothing: levels of deceit and detection in the evolution of cue-mimicry. *Proc. R. Soc. B* **286**: 20191425. http://dx.doi.org/10.1098/rspb.2019.1425

Received: 18 June 2019
Accepted: 12 August 2019

Subject Category:
Evolution

Subject Areas:
behaviour, evolution, theoretical biology

Keywords:
mimicry, deception, multi-modal signalling, costly signalling, self-deception, handicap theory

Author for correspondence:
Shahab Zareyan
e-mail: shahab.zareyan@alumni.ubc.ca

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4638155.

THE ROYAL SOCIETY PUBLISHING

A sheep in wolf's clothing: levels of deceit and detection in the evolution of cue-mimicry

Shahab Zareyan[1], Sarah P. Otto[1] and Christoph Hauert[2]

[1]Department of Zoology and Biodiversity Research Centre, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia, Canada V6T 1Z4
[2]Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, British Columbia, Canada V6T 1Z2

SZ, 0000-0002-1265-4891; SPO, 0000-0003-3042-0818; CH, 0000-0002-1239-281X

In an evolutionary context, trusted signals or cues provide individuals with the opportunity to manipulate them to their advantage by deceiving others. The deceived can then respond to the deception by either ignoring the signals or cues or evolving means of deception–detection. If the latter happens, it can result in an arms race between deception and detection. Here, we formally analyse these possibilities in the context of cue-mimicry in prey–predator interactions. We demonstrate that two extrinsic parameters control whether and for how long an arms race continues: the benefits of deception, and the cost of ignoring signals and cues and having an indiscriminate response. As long as the cost of new forms of deception is less than its benefits and the cost of new forms of detection is less than the cost of an indiscriminate response, an arms race results in the perpetual evolution of better forms of detection and deception. When novel forms of deception or detection become too costly to evolve, the population settles on a polymorphic equilibrium involving multiple strategies of deception and honesty, and multiple strategies of detection and trust.

1. Introduction

Organisms have evolved elaborate and sophisticated techniques to extract information from their conspecifics and other species [1,2]. The complexity of many of these techniques is partly owing to the prevalence of manipulated or false information, stemming from partial overlaps of interest between two interacting partners. Zahavi's handicap principle suggests that such interactions are immune to deception as long as the cost of information manipulation is high [3]. When this does not hold, however, cheating invades and potentially sets off antagonistic co-evolution between deception and detection of that deception: selection could, for example, favour prey that display to predators traits of strength and high escape capability even if they are truly undefended. In response, predators that are better able to discern weak prey are favoured, which in turn selects for prey that better hide their susceptibility, resulting in an arms race [4,5].

Implicit in the notion that better discrimination evolves is the assumption that manipulated cues are poorly coordinated with other cues that indicate the true status of an organism (e.g. defended or undefended). Hence, a more integrated level of deception is possible if these other cues evolve and become consistent with one another. This, however, is not always possible: physico-chemical and developmental factors, for example, constrain trait co-variance, making certain alterations, such as the ones required for attaining a consistent pattern of trait display, very costly or impossible. This is particularly true in the context of an arms race: although there is no reason to assume that constraints are at play initially, as the arms race proceeds, the number of cues detected increases, and with it the likelihood of constraints. In such cases, successful deception through a consistent display of traits would be expected to be costly.

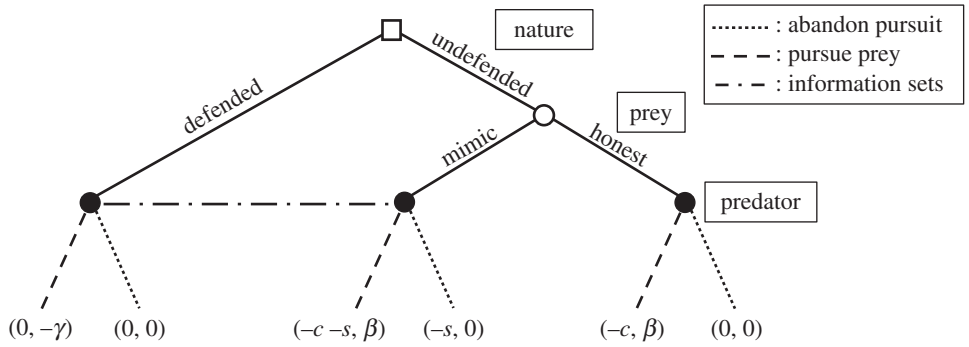

Figure 1. Game 1 in extensive form. The fitness consequence to the prey and predator are given by the first and second terms in parentheses, respectively.

We seek to determine how these well-integrated or deep forms of deception—defined more formally as deception involving manipulation of multiple primary and secondary cues that otherwise give away the deceit—evolve and are maintained in natural systems of mimicry.

In nature, cases of multi-modal deception, spanning multiple domains (ex. morphology, behaviour) and manifested through large-scale broad-acting developmental changes, provide the most convincing evidence for stable maintenance of such deep deceivers. Clearest examples of these are: (i) female mimicry in a variety of animal species, including insects, fishes, birds and mammals, initiated, in some cases, early on during development [6]; (ii) batesian mimicry in certain butterfly species, regulated by major developmental transcription factors [7–9]; (iii) morphological and behavioural mimicry of ants by spiders to evade predators [10], which, importantly, is also associated with substantial cost such as reduction in the number of eggs laid per eggsac owing to the narrowing of the bodies in the mimics [11]; and (iv) mimicry of sticks by stick insects (*Phasmatodea*), which necessitates very thin bodies and has resulted in loss of some internal organs [12].

The aim of this work is to formulate that which is common to all of the above cases in mathematical terms. We extend recent work on the evolution of partially honest systems of communication [13–15] by deriving analytical conditions for the evolution of stable systems with multi-modal detection and well-integrated but costly forms of deception. For clarity, we discuss a prey–predator cue-mimicry system, but emphasize that many interactions—both intraspecific and interspecific—involve similar forms of deceit and detection.

2. Evolution of deception

We motivate our model with the following simple scenario: a predator's hunting success is affected by the type of prey it pursues. A prey that is strong, fast, and well-armed is unlikely to be caught; the predator wastes energy attempting to pursue such a prey, and it can also sustain injuries if pursuit results in confrontation. However, pursuit of undefended prey, which are by definition slow and physically incapable of fighting back, is more likely to result in capture. As a result, it is beneficial for predators to distinguish between different prey types and pursue only the undefended.

For simplicity, we thus assume that prey fall into two categories, *defended* and *undefended*, which we assume is exclusively a function of environmental factors such as prenatal or early-life conditions. The defended corresponds to those that received adequate care and nutrition and the undefended to the ones that did not. We further assume that at any point in time a constant proportion p of prey are undefended, and the rest are defended $(1-p)$.

We analyse a single-shot game. The predator has the option of either pursuing prey or not. If pursued and the prey is defended, the predator pays a cost γ, and if the prey is undefended, the predator gains a benefit β. We assume that pursuing prey is always more beneficial than not pursuing any prey at all, that is:

$$p\beta - (1-p)\gamma > 0. \tag{2.1}$$

On the other hand, if an undefended prey is pursued, it pays a cost c. Moreover, we assume there are discernable phenotypic differences between the defended and the undefended, and that the predator has the option of detecting and basing its pursuit decision on a cue of defendedness. In turn, the undefended prey has the option of mimicking that specific cue at a cost s, in a way that prevents the predator from being able to distinguish between the two types. The game structure, which resembles the canonical Beer-Quiche signalling game used in economics [16], is presented in figure 1. In the next two sections, we consider the subsequent evolution of deception–detection and more complex forms of deception.

Six strategies can be derived from the extensive form; two for prey: (i) honest (H), corresponding to prey in their natural state and (ii) mimic (M), corresponding to prey that mimic the defended when undefended; and four for predators: (i) trusting (T), those that only pursue prey that have the cue of undefendedness and hence trust that the cue is reliable, (ii) indiscriminate (I), those that pursue prey always, (iii) perverse, those that pursue only prey that have the cue of defendedness, and (iv) predators that never pursue. Here, we focus only on the trusting and indiscriminate strategies because the others turn out to be non-essential for the dynamics, that is, the results remain qualitatively the same with or without them (see the electronic supplementary material).

The pay-offs of the two predator strategies T (Π_T) and I (Π_I) in a mixed population of honest and deceptive prey with frequencies y_H and y_D are:

$$\Pi_T = y_H p\beta \tag{2.2a}$$

and

$$\Pi_I = p\beta - (1-p)\gamma. \tag{2.2b}$$

Similarly, the pay-off of the two prey strategies H (P_H) and M (P_M) against a mixed population of trusting and indiscriminate predators with frequencies x_T and x_I are:

$$P_H = -x_I pc - x_T pc = -pc \tag{2.3a}$$

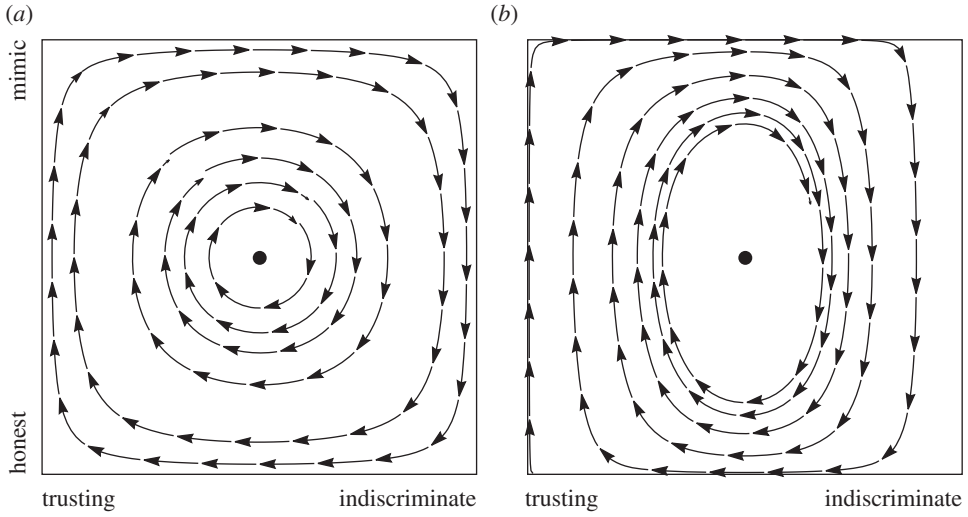

Figure 2. Surface dynamics around the (H, M; T, I) equilibrium under the (*a*) standard and the (*b*) adjusted replicator dynamics. For both panels, $p\beta = 2$, $(1 - p)\gamma = 1$, $pc = 2$ and $ps = 1$. Equilibria are indicated by block dots.

and

$$P_M = -x_I(pc + ps) - x_T ps = -x_I pc - ps. \tag{2.3b}$$

With pay-offs defined for the four strategies, we can determine how evolution unfolds using the standard replicator dynamics:

$$\dot{x}_k = x_k(\Pi_k(y) - \bar{\Pi}) \tag{2.4a}$$

and

$$\dot{y}_k = y_K(P_k(x) - \bar{P}), \tag{2.4b}$$

where x_K (y_K) is the frequency of predator (prey) strategy k, the dot denotes the time derivative, and x (y) is a vector that contains frequencies of all the predator (prey) strategies. The pay-off of predator (prey) strategy k, that is $\Pi_k(y)$ (or $P_k(x)$ for prey), is a function of the composition of the prey (predator) population, y (x). $\bar{\Pi}$ (or $\bar{P}$ for prey) corresponds to the average pay-off of strategies in the predator (prey) population.

In order to study the arms race, we assume an initial, ancestral state with honest prey and indiscriminate predators, which we denote by (H; I). If we allow new strategies to appear and fix (or be lost) before the next strategy appears, then, when deception has a net benefit against trust:

$$s < c, \tag{2.5}$$

the system keeps cycling between four states: (H; I), (H; T), (M; T) and (M; I) (see the *Mathematica* notebook for detailed calculations). Thus, evolution of cue-detection selects for cue-deception, which selects for ignoring the cue and indiscriminate pursuit, which makes cue-deception unprofitable.

On the other hand, when $s > c$ we remain at the state of honesty and trust (H; T). This latter condition, which corresponds to a case where cue manipulation is too costly, is an example of Zahavi's handicap principle [3,17].

Second, if we analyse all four strategies simultaneously, we find that there exists an interior equilibrium with non-vanishing frequencies for all four strategies (H, M; T, I), with $(y_H^*, y_M^*; x_T^*, x_I^*) = (1 - (1 - p)\gamma/p\beta, (1 - p)\gamma/p\beta; s/c, 1 - s/c)$, with the asterisks representing equilibrium frequencies (see the electronic supplementary material). This hybrid equilibrium, which is related to the hybrid equilibrium reported in recent work on signalling [13,14] (though not exactly the same as this is not strictly a signalling game), exists when condition (2.5) is met. Moreover, this equilibrium is a centre surrounded by closed orbits (figure 2*a*) [18]. The direction of movement around the centre mimics the cyclical invasions between the four monomorphic states discussed above.

As noted by Maynard Smith [19], these closed orbits are sensitive to the assumptions made when deriving the replicator dynamics [20]. For a slightly different version of these dynamics, see the adjusted replicator dynamics [19,21], which are exactly the same as the standard dynamics with exactly the same equilibria except that the fitness values in the adjusted dynamics are divided by mean fitness (equations (2.6)), the cycles contract and ultimately converge to the interior equilibrium (figure 2*b*):

$$\dot{x}_k = x_k \frac{\Pi_k(y) - \bar{\Pi}}{\bar{\Pi}} \tag{2.6a}$$

and

$$\dot{y}_k = y_k \frac{P_k(x) - \bar{P}}{\bar{P}}. \tag{2.6b}$$

As can be seen from equations (2.6), the adjusted replicator dynamics is a consequence of a simple change in assumptions about how selection acts on a population: if the mean fitness of a population is low, selection is stronger and change happens more rapidly under the adjusted regime as compared to the standard dynamics. Other processes, including mutation [15], can also lead to the asymptotic stability of the hybrid equilibrium. Thus, under slight and reasonable changes to the underlying assumptions, the equilibrium with all the strategies becomes stable.

Why is this interior equilibrium significant? Traditionally, many have argued that the evolution of deception does not necessarily lead to deception–detection. Instead, the deceived can simply ignore the cue [2]. This is indeed true, as is shown in the homoclinic cycle connecting the four monomorphic equilibria discussed above. However, the hybrid equilibrium (whether stable or surrounded by closed orbits) demonstrates that *ignoring the cue* and *detecting deception* are not two mutually exclusive possibilities. Indeed the hybrid equilibrium includes indiscriminate predators that ignore all cues

and trusting predators that are *always* deceived by the mimic. The stable existence of deception should then allow more sophisticated means of deception–detection to evolve.

3. Evolution of deception–detection

Deception through the trait considered above (i.e. that used by the mimic) might cause physiological, anatomical, or behavioural changes that give away the deceit: if, for instance, the length of legs was originally detected to differentiate between defended and undefended prey, then the undefended mimic could evolve longer legs to deceive the predators. An unintended consequence of this, however, is that leg length changes relative to other traits, and this asymmetry can then be detected as a secondary cue that gives away the deceit. Hence, prior to deciding whether to pursue a prey, we give predators the option of detecting a secondary cue at a cost δ to reveal the deception and distinguish the mimic from the model (i.e. the type that is being mimicked, in this case, the defended prey).

Allowing for this option results in 12 strategies for the predator (see the electronic supplementary material for the extensive form), but only three are essential for the dynamics (for details see the electronic supplementary material): (i) trusting, (ii) indiscriminate and (iii) detector (D), that is, predators that detect the secondary cue and pursue both honest and mimetic undefended prey. The detector's pay-off against a mixed population of honest and deceptive prey is:

$$\Pi_D = y_H(p\beta - \delta) + y_M(p\beta - \delta) = p\beta - \delta. \tag{3.1}$$

Comparing equation (3.1) with equations (2.2), we realize detectors have a higher pay-off than trusting predators if

$$\delta < y_M p\beta, \tag{3.2}$$

that is, if the cost of improved detection is less than the weighted benefits of reducing false negatives (not pursuing the mimetic undefended). Moreover, detectors have a higher pay-off than indiscriminate predators if

$$\delta < (1-p)\gamma, \tag{3.3}$$

that is, if the cost of improved detection is less than the weighted cost of false positives (pursuing the defended).

Detectors invade the hybrid equilibrium (H, M; T, I) whenever condition (3.3) is met. Thus, the (H, M; T, I) hybrid equilibrium represents an evolutionary endpoint only if improved detection is too costly. In order to analyse the consequences of invasion, we first adjust the pay-offs of the honest and the deceptive prey to account for the new detecting predators with frequency x_D:

$$P_H = -x_I pc - x_T pc - x_D pc = -pc \tag{3.4a}$$

and

$$P_M = -x_I p(s+c) - x_T ps - x_D p(s+c) = -(x_I + x_D)pc - ps. \tag{3.4b}$$

If detectors invade, an equilibrium analysis with all the strategies shows that the only locally stable equilibrium is one composed of honest and deceptive prey, alongside trusting and detector predators (H, M; T, D). Given that this equilibrium exhibits asymptotic stability under the adjusted dynamics [18] (and neutral stability under the standard dynamics), it is as robust as the original hybrid equilibrium. Overall, the results here imply that costly detection cannot evolve if false-positive errors (mistaking a defended prey as undefended) are of low cost (low γ) or if there are very few defended models in the population (high p).

At the new (H, M; T, D) equilibrium, $(y_H^*, y_M^*; x_T^*, x_D^*) = (1 - \delta/p\beta, \delta/p\beta; s/c, 1 - s/c)$. The frequency of mimics $\delta/p\beta$ is lower than their frequency at the previous equilibrium $((1-p)\gamma/p\beta)$ because equation (3.3) must hold for detectors to spread. In other words, the original hybrid's prey population evolves towards more honesty upon evolution of deception–detection.

4. Evolution of well-integrated deception

The evolution of detection creates directional selection on the prey to better deceive detectors by manipulating secondary cues. This more integrated deception necessitates concealment of not only the undefended state of the prey but also the deception itself. In contrast to simple deception, it requires the decoupling of not only the primary trait from the state of the organism, but also cues of deception from deception itself. For this reason, more sophisticated, or well-integrated levels of deception, are more likely to require costly broad-acting changes compared to adaptations involving manipulation of a single trait.

To study the evolution of well-integrated deception, we allow prey that mimic the defended in the primary cue the ability to mimic the defended in the secondary cue as well, at a cumulative cost g, which we naturally assume to be greater than the cost of simple deception ($g > s$) (for the extensive form of the interaction, see the electronic supplementary material). The game results in three strategies for prey and 12 strategies for predator. In the electronic supplementary material, we demonstrate that the strategies that matter for the evolutionary dynamics are the same as the ones discussed above plus the additional strategy of well-integrated deception (W), which corresponds to prey that mimic the defended in both the primary and the secondary cues. The pay-off of this strategy is:

$$P_W = -x_I p(c+g) - x_T pg - x_D pg = -x_I pc - pg. \tag{4.1}$$

Comparing equations (3.4) and (4.1) shows that in a population of trusting and detecting predators, well-integrated deception has a higher pay-off than simple deception if

$$g - s < x_D c, \tag{4.2}$$

or, in other words, when the residual cost of well-integrated deception ($g - s$) is less than the cost of being pursued when undefended, weighted by the frequency of detectors. This result implies that more costly forms of well-integrated deception (higher g) can invade simple mimicry when detectors are more common. Furthermore, well-integrated deception has a higher pay-off than honest prey against trusting and detector predators if

$$g < c, \tag{4.3}$$

that is, if the overall cost of well-integrated deception does not exceed the costs of being pursued. Importantly, the condition for well-integrated deceivers to invade the hybrid equilibrium (H, M; T, D) is also $g < c$. This means that the evolution of well-integrated deception does not depend on the costs of simple mimicry. Well-integrated morphs invade the hybrid equilibrium above as long as they deceive the undeceived and as long as their cost is

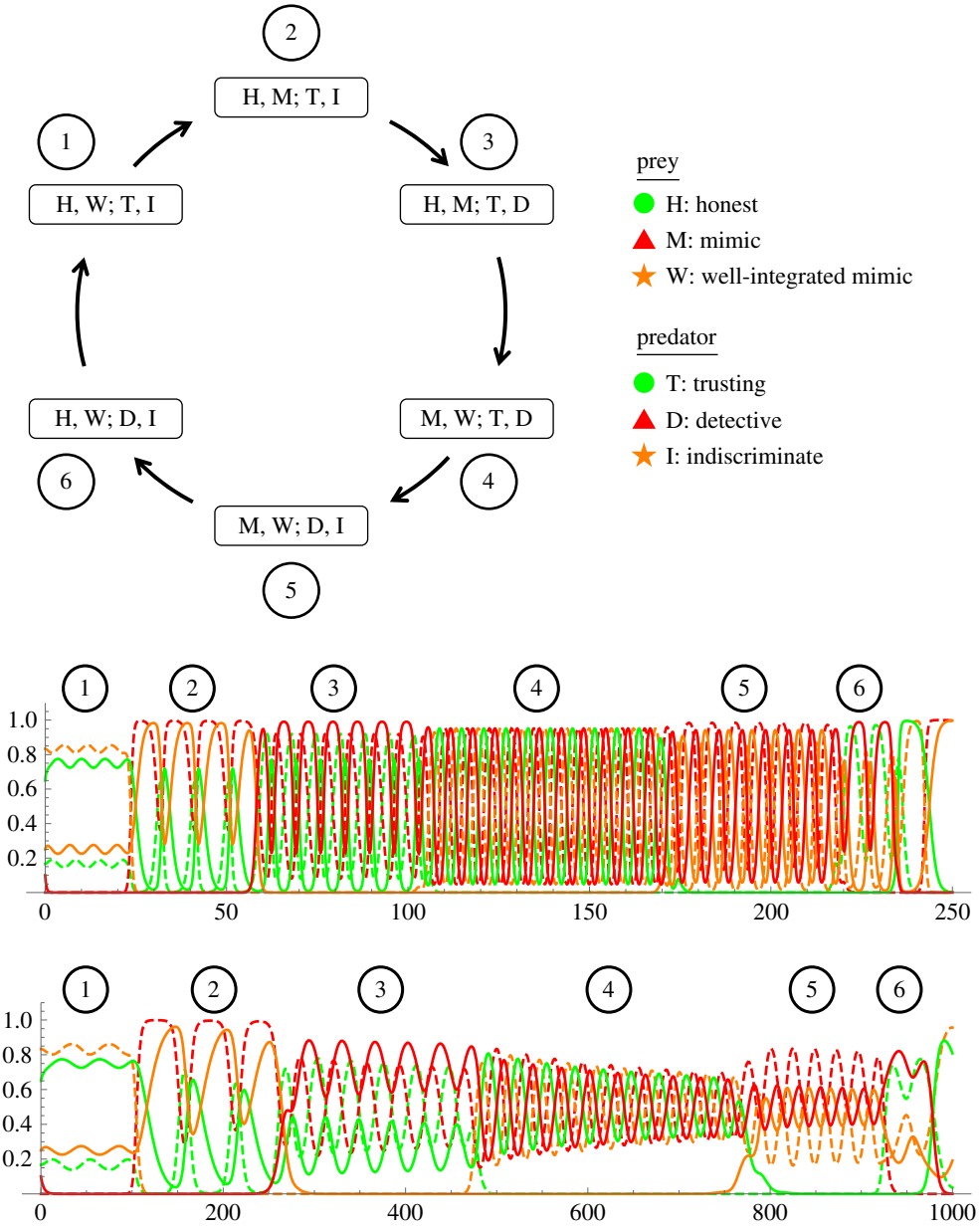

Figure 3. Cycles of invasions and re-invasions upon introduction of well-integrated mimics. Top: schematic of the cycle. The rationale behind each jump is explained in the electronic supplementary material. Bottom: movement between the six equilibria in real time for standard (top) and adjusted (bottom) dynamics. Dashed lines are prey, solid lines are predators. The equations were solved numerically for a brief period of time with a very low mutation rate of 10^{-30} (to prevent immediate convergence to the interior tri-morphic equilibrium), starting at the (H, W; T, I) equilibrium. Parameters: $pc = 6$, $pg = 4.5$, $ps = 1.5$, $p\beta = 6$, $(1-p)\gamma = 5$, $\delta = 3$. (Online version in colour.)

less than c. In this model, c is the cost of being pursued when undefended. More generally, c can be re-interpreted as the benefits of successful deception. Hence, the result here implies that if benefits of more integrated deception are too low, well-integrated deception does not evolve. Stated in a different way, if there are large benefits to successful deception, well-integrated deception evolves in our model even if it engenders high cost.

In order to analyse the invasion of well-integrated deception with frequency y_W, we first modify the pay-offs of the three predator strategic types:

$$\Pi_I = p\beta - (1-p)\gamma, \tag{4.4a}$$

$$\Pi_T = y_H p\beta \text{ and} \tag{4.4b}$$

$$\Pi_D = y_H(p\beta - \delta) + y_M(p\beta - \delta) - y_W\delta = (y_H + y_M)p\beta - \delta. \tag{4.4c}$$

Analysis of the replicator equations shows that, following the invasion, well-integrated deceivers replace honest prey, and the populations move towards an (M, W; T, D) equilibrium. Thus, the prey population is composed only of deceivers, and as a result, the trusting predators are always subjected to deception: they never pursue any prey and their pay-off is zero.

What will happen next? If only one strategy is introduced at a time, the system shifts from one hybrid equilibrium to another in an endless cycle (see figure 3; electronic supplementary material). If, however, we allow all strategic types to occur, a central hybrid equilibrium exists, composed of honesty, normal deception and well-integrated deception in the prey population, and trusting, detector and indiscriminate in the predator population (H, M, W; T, D, I), with $(y_H^*, y_M^*, y_W^*; x_T^*, x_D^*, x_I^*) = (1-(1-p)\gamma/p\beta,\ \delta/p\beta,\ ((1-p)\gamma-\delta)/p\beta;\ s/c,\ (g-s)/c,\ 1-g/c)$. The tri-morphic equilibrium of the replicator dynamics is again a neutrally stable centre (see

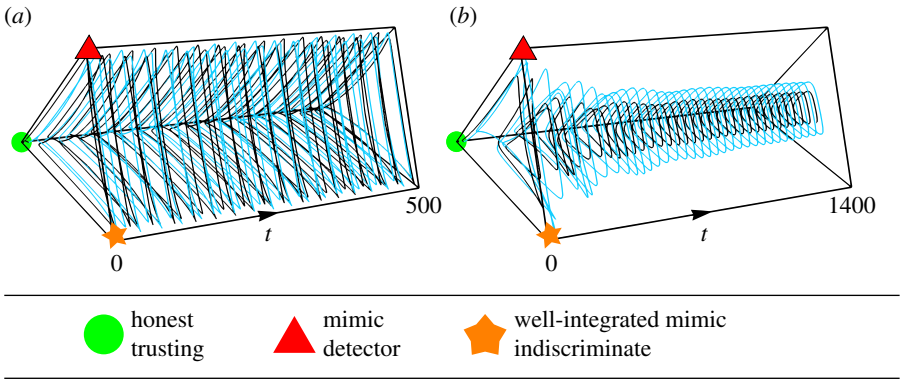

Figure 4. Dynamics of prey (blue/grey) and predators (black) around the (H, M, W; T, D, I) equilibrium point under the (*a*) standard and (*b*) adjusted replicator dynamics. The time period over which the replicator equations were numerically solved is indicated at the bottom of each subfigure. For both panels, t indicates time, $p\beta = 3$, $pc = 3$, $ps = 1$, $(1-p)\gamma = 2$, $\delta = 1$ and $pg = 2$. (Online version in colour.)

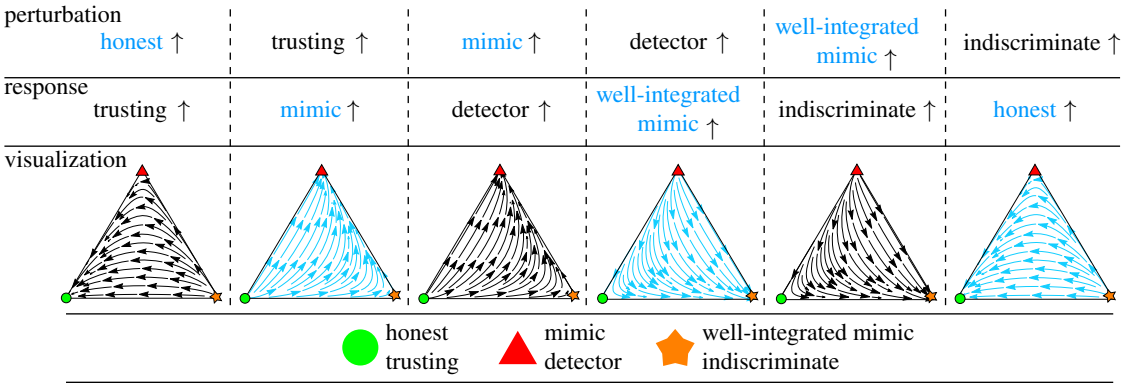

Figure 5. Response to perturbations away from the tri-morphic equilibrium in prey (blue/grey) and predator (black) populations. For all panels, we perturb the system in directions indicated at the top of the figure. We then determine the response to each perturbation using the standard replicator dynamics (second row), and visualize the response by drawing vector fields that depict selection gradients following each perturbation. Note that perturbing the system in one of the populations (say, prey) induces selection only in the other population (say, predator)—not in the original population. Parameters: $p\beta = 3$, $(1-p)\gamma = 2$, $\delta = 1$, $pc = 3$, $ps = 1$ and $pg = 2$. (Online version in colour.)

the *Mathematica* notebook). Numerical analysis of the dynamics under the adjusted replicator equation shows convergence to this single equilibrium point (figure 4). A few aspects of this equilibrium are noteworthy.

First, to develop a better understanding of why coexistence is possible, we visualize the interactions between the various strategies by plotting the vector fields that represent the selection dynamics (figure 5). As can be seen, the antagonistic interactions between pairs of strategies prevent any one strategy from driving another extinct.

Second, the frequency of undetected deception is high if false positives are costly (high γ) and if there are many defended models in the population (low p). This is seen both at the above equilibrium of (H, M, W; T, D, I), where the frequency of well-integrated deceivers is $((1-p)\gamma - \delta)/p\beta$, and at the original hybrid equilibrium of (H, M; T, I), where undetected deception has frequency $(1-p)\gamma/p\beta$.

Third, although the total frequency of M and W deceivers at this (H, M, W; T, D, I) equilibrium is the same as their frequency at the original (H, M; T, I) hybrid equilibrium ($(1-p)\gamma/p\beta$), the relative proportion of the two deceptive strategies is determined by δ, the cost of detecting simple forms of deception. If detection has low cost, well-integrated deception would be more frequent than simple deception.

Lastly, if well-integrated deception is very costly relative to simple deception (high $g - s$), our model shows that the equilibrium frequency of detectors would be high. Overall, if γ and g are high, but δ and s are low, well-integrated deceivers and fine-tuned detectors will come to constitute the majority of the population.

5. Discussion

This study emphasizes that, even with a low cost of cue manipulation, the spread of deception does not necessarily result in the collapse of cue-detection or information transfer in general. We find that deception can coexist indefinitely in a heterogeneous population alongside honesty, in line with [13]. The mechanism that allows for this heterogeneity is a simple one: once common, deception is counter-selected by the spread of strategies that ignore the mimicry, which leads to the stable coexistence of multiple strategies in both populations. Given the simplicity of the mechanism, we expect it to apply to many systems of communication in nature and expect them to be heterogeneous [13].

Although previous work had predicted that heterogeneous signalling systems are likely to include types that never respond to signals [13], here we demonstrate that expanding the options available to predators opens up new possibilities: prolonged existence of deception in a population can result in selection for those mutants that detect and avoid

the deception. Thus, besides the presence of constitutive responders, it is also likely for detectors to evolve the ability to detect multiple cues or seek multiple sources of information. Deception is nevertheless maintained in the population if detection is costly and only favoured when deceivers are common, and deceivers are common only when detectors are rare: neither can drive the other extinct.

Importantly, we show that, as a consequence of the evolution of multi-modal detectors, mimicry can evolve to be well-integrated, possibly involving broad-acting and costly changes that allow for the decoupling of cues of deception from the deception itself. Certainly, the many forms of deception in nature that are manifested through major developmental changes are consistent with this conclusion. Indeed, it is conceivable for the cost of deception owing to broad-acting changes (g) to be so high that its net potential benefit ($c - g$) approaches zero, and yet such complex and well-integrated deception is maintained stably in the population. Thus, high costs should not be taken as necessarily providing support for the costly signalling theory, as complex forms of deception can invade systems with simpler forms of deception, raising the apparent costs. Most importantly, the analysis demonstrates that the properties of the heterogeneous equilibria are governed by several key parameters. For example, if false positives are of low cost and/or if the model/mimic proportion is low (equation (3.3)), then detection of high cost cannot evolve, and hence, well-integrated deception cannot evolve either. Also, if deception has a small benefit (equation (4.3)), then well-integrated deception, which we assume is of high cost, is unlikely to evolve. Moreover, the frequency of well-integrated deceivers at the hybrid equilibrium depends on the cost of false positives (attacking defended prey), benefit of true positives (attacking undefended prey), model/mimic ratio and cost of detection. All of this can be translated into specific predictions when studying various natural systems. Overall, these results demonstrate that the properties of the arms race, such as whether costly and complex deception and detection can evolve and thus how long the arms race continues, are heavily regulated by extrinsic parameters that are specific to each system (ex. benefit of successful mimicry, cost of false positives, etc.).

In this work, we assumed that the defended prey does not evolve. It is, however, conceivable that following the evolution of deceivers, the defended prey evolve to exaggerate the detected trait (the cue) in order to distinguish themselves better from the mimics. Exaggerated traits would be more costly to mimic (higher s) and as a consequence, this could result in the evolution of more honest populations. We also assumed that the rate of predation remains constant over the time course of the evolutionary game being analysed, effectively assuming that the density of predators is not controlled by the focal species and that predators do not become satiated if they succeed more often in capturing this prey species. Alternative models worth exploring in the future would allow for a numerical response in the predators and for satiation following successful predation events, which would generate further density- and frequency-dependent interactions. Consequently, predators' 'motivation' to distinguish between different kinds of prey would change, causing the cost and benefit structure to vary over time.

The results of this work are relevant for a variety of cue-detection systems, some of which were discussed in the introduction. They also can be interpreted in the context of signalling and communication; the interaction analysed in the first section with (H, M, T, I) is a modified version of the canonical Beer-Quiche signalling game in economics [16]. In human communication, the results are directly applicable to Trivers' hypothesis on the evolution of self-deception [22]. This hypothesis states that self-deception, defined as the deception of the conscious part of the mind by the subconscious, either through biasing the gathering of information or biasing gathered information, evolved for the better deception of others: by virtue of believing in their own lies, self-deceivers do not give secondary cues that otherwise give away the deceit (equivalent to the well-integrated strategy). Such masterful deception should be very costly, as it causes a biased perception of reality and suboptimal decision-making. Trivers hypothesizes that this could have only evolved as a result of an arms race between deception and detection. Thus many properties of self-deception as a strategy are shared with well-integrated mimicry. Trivers notes: 'It stands to reason that if our theory of self-deception rests on a theory of deception, advances in the latter will be especially valuable' [12, p. 50]. Although the study of self-deception is in infancy, here we have provided a systematic analysis of the evolution of well-integrated deception that will hopefully provide the groundwork for further study.

Data accessibility. All data are included in the electronic supplementary material.

Authors' contributions. Conceived and designed the experiments: S.Z., C.H., S.P.O. Performed the experiments: S.Z. Analysed the data: S.Z. with assistance from S.P.O., C.H. Wrote the paper: S.Z. with assistance from C.H., S.P.O. All authors gave final approval for publication.

Competing interests. We declare we have no competing interests.

Funding. Research funding was provided by the Natural Sciences and Engineering Research Council of Canada (CGSM scholarship to S.Z.; RGPIN-2016-03711 to S.P.O.; RGPIN-2015-05795 to C.H.).

Acknowledgements. We thank A. Ghaseminejad for extensive comments and discussion on the earlier versions of the manuscript, T. P. Flower and M. M. Osmond for feedback on the entire paper, as well as members of the Otto, Hauert and Doebeli laboratories for feedback on the project. We also thank Carl Bergstrom for deeply insightful suggestions that formalized the analysis in the manuscript, as well as the recommendations of an anonymous reviewer that helped us make the language of this project more precise and accessible.

References

1. Taga ME, Bassler BL. 2003 Chemical communication among bacteria. *Proc. Natl Acad. Sci. USA* **100**(Suppl. 2), 14 549–14 554. (doi:10.1073/pnas.1934514100)
2. Searcy WA, Nowicki S. 2005 *The evolution of animal communication: reliability and deception in signaling systems*. Princeton, NJ: Princeton University Press.
3. Zahavi A. 1975 Mate selection: a selection for a handicap. *J. Theor. Biol.* **53**, 205–214. (doi:10.1016/0022-5193(75)90111-3)
4. Dawkins R, Krebs JR. 1979 Arms races between and within species. *Proc. R. Soc. B* **205**, 489–511. (doi:10.1098/rspb.1979.0081)
5. Krebs JR, Dawkins R. 1984 Animal signals: mind-reading and manipulation. In *Behavioral ecology: an evolutionary*

approach, 2nd edn (eds JR Krebs, NB Davies), pp. 380–403. Oxford, UK: Blackwell Scientific Publications.
6. Saetre GP, Slagsvold T. 1996 The significance of female mimicry in male contests. *Am. Nat.* **147**, 981–995. (doi:10.1086/285889)
7. Kunte K, Zhang W, Tenger-Trolander A, Palmer D, Martin A, Reed R, Mullen SP, Kronforst MR. 2014 Doublesex is a mimicry supergene. *Nature* **507**, 229–232. (doi:10.1038/nature13112)
8. Nishikawa H. 2015 A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat. Genet.* **47**, 405–409. (doi:10.1038/ng.3241)
9. Timmermans MJTN. 2014 Comparative genomics of the mimicry switch in *Papilio dardanus*. *Proc. R. Soc. B* **281**, 20140465. (doi:10.1098/rspb.2014.0465)
10. Shamble PS, Hoy RR, Cohen I, Beatus T. 2017 Walking like an ant: a quantitative and experimental approach to understanding locomotor mimicry in the jumping spider *Myrmarachne formicaria*. *Proc. R. Soc. B* **284**, 20170308. (doi:10.1098/rspb.2017.0308)
11. Cushing PE. 1997 Myrmecomorphy and myrmecophily in spiders: a review. *Fla. Entomol.* **80**, 165–193. (doi:10.2307/3495552)
12. Trivers R. 2011 *The folly of fools: the logic of deceit and self-deception in human life*. New York, NY: Basic Books.
13. Zollman KJ, Bergstrom CT, Huttegger SM. 2012 Between cheap and costly signals: the evolution of partially honest communication. *Proc. R. Soc. B* **280**, 20121878. (doi:10.1098/rspb.2012.1878)
14. Huttegger SM, Zollman KJ. 2010 Dynamic stability and basins of attraction in the Sir Philip Sidney game. *Proc. R. Soc. B* **277**, 1915–1922. (doi:10.1098/rspb.2009.2105)
15. Huttegger SM, Zollman KJ. 2016 The robustness of hybrid equilibria in costly signaling games. *Dyn. Games Appl.* **6**, 347–358. (doi:10.1007/s13235-015-0159-x)
16. Cho IK, Kreps DM. 1987 Signaling games and stable equilibria. *Q. J. Econ.* **102**, 179–221. (doi:10.2307/1885060)
17. Grafen A. 1990 Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546. (doi:10.1016/S0022-5193(05)80088-8)
18. Hofbauer J, Sigmund K. 1998 *Evolutionary games and population dynamics*. New York, NY: Cambridge University Press.
19. Smith JM. 1988 *Evolution and the theory of games*. Berlin, Germany: Springer.
20. Traulsen A, Claussen JC, Hauert C. 2005 Coevolutionary dynamics: from finite to infinite populations. *Phys. Rev. Lett.* **95**, 238701. (doi:10.1103/PhysRevLett.95.238701)
21. Weibull JW. 1997 *Evolutionary game theory*. Cambridge, MA: MIT Press.
22. Trivers R. 1985 *Social evolution*. Menlo Park, CA: Benjamin/Cummins Publishing Co.