

## Using quantitative methods of determining group membership to draw biological conclusions



Steven V. Viscido<sup>\*</sup>, Shubha Shrestha

Department of Biological Sciences, Winston-Salem State University, Winston-Salem, NC, U.S.A.

### ARTICLE INFO

#### Article history:

Received 15 December 2014  
 Initial acceptance 12 January 2015  
 Final acceptance 9 February 2015  
 Available online 13 April 2015  
 MS. number: AS-14-01019R

#### Keywords:

aggregation index  
 ambit  
 animal group  
 modelling  
 simulation

For many years, biologists have agreed upon a clear definition of a population: all the members of a single species that exist in a given habitat at the same time and that at least potentially interbreed (Barbour, Burk, & Pitts, 1987; Ricklefs & Relyea, 2014). Thus, even though plants may be hundreds of metres apart, as long as there is the possibility of genetic exchange (e.g. via wind pollination) they are normally judged to be part of the same population (Barbour et al., 1987). This universally recognized definition ensures that population biologists use a common language and enables investigators to draw conclusions across a broad range of taxa, leading to valuable ecological insights.

Sadly, no such commonly accepted definition has ever been provided for an animal group. To be sure, 'classes' of animal gatherings have been rigorously defined, such as the fish school, defined as a polarized (i.e. closely aligned) group (Shaw, 1978). However, such definitions rely on the investigator having already classified a set of animals as a group. When it comes to defining a group itself, investigators seem to make such determinations using criteria that, all too frequently, do not agree with those of other investigators (Parrish, Viscido, & Grünbaum, 2002). Indeed, when we examine studies of group behaviour, we find three broad categories of studies: (1) those providing a generalized, but unquantifiable,

definition of the group; (2) those providing a rigorously quantified, but idiosyncratic (i.e. particular to a given study), definition of the group; and (3) those that refer to a 'group' (or 'flock', 'herd', 'aggregation', etc.) without explicitly defining the term.

Review papers and large integrative studies frequently express a broad, general definition of the group. For example, Williams (1964, page 353) defined aggregations as 'groups that arise by individuals independently seeking the same localized conditions.' Pitcher and Parrish (1993, page 365) defined groups as animals 'that remain together for social reasons'. Ritz (1994, page 158), following Wilson (1975), defined social aggregations as the 'mutual attraction of conspecific animals leading to co-operative group behaviour'. Although we certainly do not disagree with these generalized definitions, we cannot discover any objective means by which we could, upon observing a collection of individuals in the field (or the laboratory), use these definitions to determine whether a set of animals in arbitrary positions are all members of a single group. These broad definitions thus defy our ability to quantitatively characterize the group.

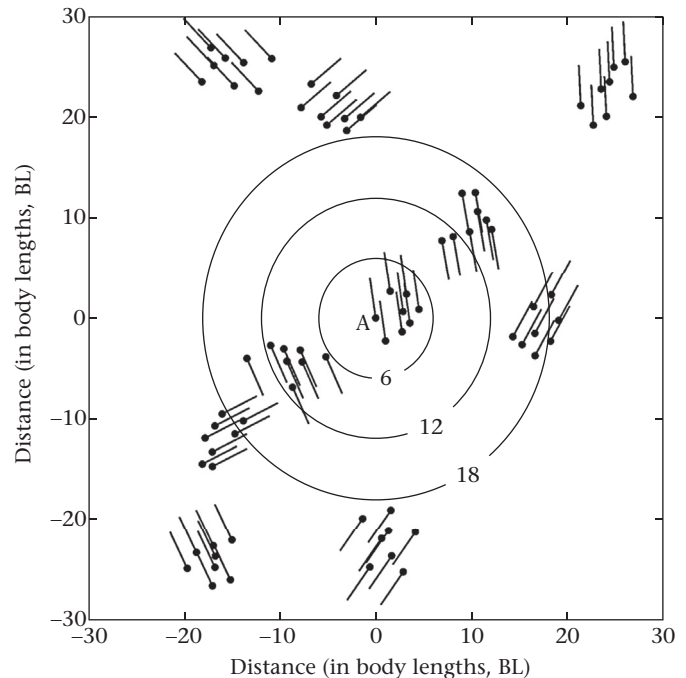
Less general studies of specific organisms have provided a rigorously quantified definition of the group, but their definitions have not been consistent. For example, in the field, Heard (1992) defined musk-oxen as being group-mates if they were within 'a few meters' of each other, as compared with intergroup distances, which were much greater (kilometres). In laboratory experiments, Pitcher and Parrish (1993) defined individuals within four body lengths (BL) of each other as group-mates. On the other hand, grid-

<sup>\*</sup> Correspondence: S. V. Viscido, Department of Biological Sciences, Winston-Salem State University, Winston-Salem, NC 27110, U.S.A.  
 E-mail address: [viscidost@wssu.edu](mailto:viscidost@wssu.edu) (S. V. Viscido).

based cellular automata (CA) models often define groups as being separated by at least one empty row of grid squares (Stöcker, 1999; Vabø & Nøttestad, 1997). In contrast to CA models, in individual-based simulations, Flierl, Grünbaum, Levin, and Olson (1999) and Gautrais, Jost, and Theraulaz (2008) defined the group as all the individuals within the sensing range of at least one other group member (i.e. those connected to one another either directly or indirectly). These definitions of the group are biologically meaningful and mathematically tractable, but because each study provides its own algorithm for defining a group, the methods used among different studies do not agree. These differences make it difficult to compare findings across studies, thus harming our ability to draw broad biological conclusions.

As challenging as it is to compare studies that use differing definitions of a group, even more difficulty arises when the method of assigning group membership is not made explicit. In the laboratory, frequently all individuals in a confined arena (e.g. the same fish tank) are considered 'group-mates' for the purposes of the experiment (Katz, Tunström, Ioannou, Huepe, & Couzin, 2011; Salierno, Gipson, & Kane, 2008), and in such studies no other definition is needed. On the other hand, in field studies without spatial confinement, the authors must determine groups using some other set of criteria. For example, Ballerini et al. (2008) used a three-dimensional alpha-shape algorithm to detect group edges of bird flocks. Unfortunately, authors sometimes do not explicitly state their criteria for assigning group membership. For example, Petit and Bildstein (1987) reported the occurrence of solitary, paired and grouped white ibises, *Eudocimus albus*, in a South Carolina marsh, but they provided no insight into what criteria they used to differentiate these categories. Similarly, Wrona and Dixon (1991) reported the number of pupal groups per rock and the number of pupae per group in trichopterans (*Ryacophila vao*), but did not provide the metrics by which they made these determinations. And in simulation, several studies have included the ability for groups to split or merge, but do not provide the algorithm used to determine group membership (e.g. Cambui & Rosas, 2012; Zheng, Kashimori, Hoshino, Fujita, & Kambara, 2005). In such cases, one is left wondering how groups were identified. Additionally, because it is unlikely that similar algorithms were used by all these studies, the results cannot be easily compared, which, again, decreases our ability to draw biological insights.

The lack of a consensus about how one should identify individuals as group members is a serious issue, because how one defines 'a group' can dramatically affect the biological conclusions we draw from the data. For example, let us imagine a large habitat within which 10 small groups are scattered across the field (Fig. 1). Let the group size be  $G = 8$  individuals per group. Let the individuals within each group exhibit allelomimesis, resulting in perfect alignment  $\phi$  within each group as defined by Viscido, Parrish, and Grünbaum (2007) ( $0^\circ$  angle of deviation gives  $\phi = 1$ ;  $180^\circ$  of angle deviation gives  $\phi = 0$ ). Additionally, let individuals ignore the members of other groups, so that each group is positioned and oriented randomly with respect to the other groups. If one centres the field of view on one individual (marked 'A' in Fig. 1), one can properly identify that individual's group-mates by drawing a circle of the correct size around it, and counting those within the circle (Fig. 1, radius = six body lengths (BL)). Assigning group membership in this fashion would lead to correct estimates of both group size  $G_{\text{est}}$  and alignment  $\phi_{\text{est}}$  ( $G_{\text{est}} = 8$ ,  $\phi_{\text{est}} = 1$ ). However, if we were to (incorrectly) define the interaction circle as being larger (Fig. 1, radii = 12 BL or 18 BL), we would arrive at the wrong conclusions: individuals from 'other' groups would be counted, and we would overestimate group size ( $G_{\text{est}} = 22$  and 38, respectively), and underestimate alignment ( $\phi_{\text{est}} = 0.63$  and 0.61, respectively). Thus, without a proper working definition of a group, and the means to



**Figure 1.** A plot of an example population of  $I = 80$  individuals divided into 10 groups of 8. Orientation of all individuals in a given group is perfectly aligned, but orientation among groups is random. Concentric circles for interaction radii of 6, 12 and 18 body lengths are shown around individual 'A'. If the group is defined using radii of 12 or 18, individuals who are not part of A's group will be incorrectly included in group metric calculations.

accurately assign group membership, we have little hope of understanding aggregation biology.

Clearly, a practical operational definition of an animal group is sorely needed. Therefore, we have two goals for this paper. First, we propose a meaningful biological definition for 'a group' that is reminiscent of the classical definition of a population. And second, we provide an example quantitative method for applying our definition to properly identify members of both real and simulated animal aggregations, including an algorithm that can easily be applied to many sets of animals in known positions. Our hope is that we can help improve the comparability of group behaviour studies by providing future investigators with both a useful definition and an example of how to apply it to their research.

## DEFINING THE GROUP

Historically, definitions of an animal group have one major factor in common: that the group members are interacting with one another. As Pitcher and Parrish (1993, page 365) put it, the group 'remain[s] together for social reasons'. These social reasons can include improved foraging success (Beauchamp, 2014), predation defence (Bauchamp, 2012; Hamilton, 1971; Zheng et al., 2005), or reproductive benefits (Burger, 1981; Ritz, 1994) that accrue to group members. These interactions can be quite varied. For example, Götmark, Winkler, and Andersson (1986) described how black-headed gulls, *Larus ridibundus*, use the foraging dives of their flock-mates to locate food. Viscido and Wethey (2002) showed that fiddler crabs, *Uca pugnator*, rapidly approach their group-mates when frightened by an attacking predator. Similarly, colonial breeding can improve the survival of chicks as compared with solitary breeding (Burger, 1981). Thus, a group is clearly made up of individuals who derive benefits from regularly interacting with one another.

On the other hand, especially in larger groups, these interactions need not all be direct, and in some cases, such as fish schools numbering in the hundreds of thousands, the animals clearly cannot all see or interact with even a fraction of the whole group (Parrish, 1991). And yet, we still consider them to be a group because the group acts like a single entity, frequently exhibiting coherent emergent properties, such as a common alignment, that result from individual interactions, but are not up to any single individual to determine (Couzin & Krause, 2003; Flierl et al., 1999; Parrish & Edelstein-Keshet, 1999). Consequently, we must include not only direct interactions, in which an individual senses and responds to a neighbouring group member, but also indirect interactions, in which the individual cannot itself sense the actions of the second animal, but is responding to neighbours who do (Fig. 2). Therefore, we can describe a group as all members of a given population who are 'currently' interacting with one another either directly or indirectly, and thus arrive at a definition of an animal group that we think is both generally applicable and, with admittedly some effort, quantifiable both for living animals and (somewhat more easily) for model systems. Following closely with several previous studies (e.g. Flierl et al., 1999; Gautrais et al., 2008; Gueron, Levin, & Rubenstein, 1996), we define the animal group as 'the set of all the members of a population who are in the interaction zone of at least one other member of the set during the observation period'. Thus, group members are like links in a chain (Jarman, 1987); or, as Gautrais et al. (2008) put it, the interaction zones of all group members must be 'connected'.

This definition requires us to know two things: (1) the positions of all animals on the field, and (2) the interaction zone of each animal. The former is far easier to measure than the latter, and clearly, for living animals, the main sticking point of this definition will be estimating the size of the interaction zone (for models, it can be defined by the investigator).

In the next section, we will demonstrate the use of positional data to infer group membership, but first, we explain the final term in our definition, the 'observation period'. Because there is so much variety both in animal behaviour across taxa and in the purposes of

each study, defining the relevant observation period must be left to each investigator. In most cases, the observation period will be a 'time step' (e.g. one time cycle of a model; one frame of video; one scan sample period), and group membership should be assessed separately for each time step. Significantly, defining the observation period as a single time step allows us to reassess group membership over time and permits the observer to discover patterns of group fission and fusion as individuals join and leave groups.

## METHODS

Having defined a group, we now turn to a straightforward method for objectively assigning group membership to a collection of individuals in known positions. We note that this method can only be used when the positions of all animals are at least approximately known. To demonstrate how the method works, we created an object-oriented MatLab program to generate simulated animal populations and to conduct statistical analysis. In the program, each simulated animal is an object, with a position and a velocity vector. These objects exist within a 'container' object, the habitat. Our model was run in two dimensions, but the same procedures can be used in three dimensions without loss of generality.

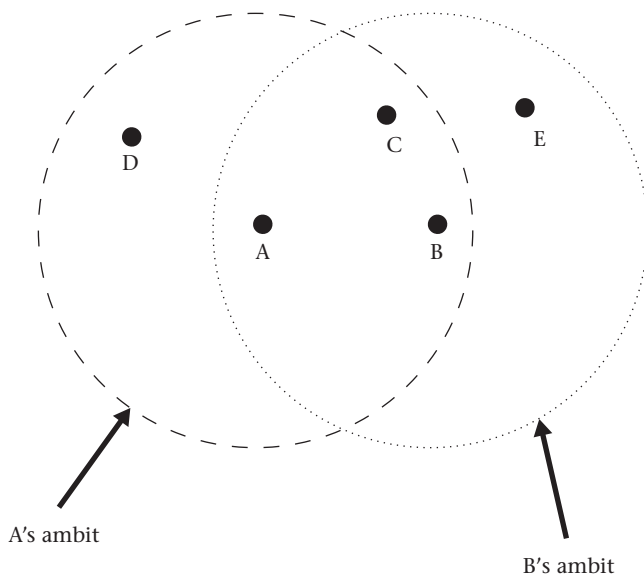
To assign each individual to its group, we used a method we call the 'ambit method' after Lloyd (1967). We define the ambit as the circular area (or a spherical area in three dimensions) drawn around a focal individual that contains all the individuals who exert a 'direct' behavioural effect on that focal animal (Lloyd, 1967). Thus, the ambit is the 'interaction zone' of the focal animal. In zone-based fish-schooling models, for example, the ambit would be the zones of repulsion, alignment and attraction, and indeed, except for minor differences in terminology, this algorithm has been used before to determine group membership (e.g. Flierl et al., 1999; Gautrais et al., 2008; Gueron et al., 1996). Continuing to follow Lloyd (1967), if we let  $a$  represent the circular ambit area (say, in square BL), then we can compute the interaction distance  $D$  as the circle's radius:

$$D = \sqrt{\frac{a}{\pi}}. \quad (1)$$

Thus, if we know the position of each animal on the plane, then by computing distances among them, we can discover who is within an interaction distance  $D$  to each individual, which provides us with a list of the 'ambit-mates' for each animal. Because of how we have defined group membership, we can say that 'an individual is in the same group with all of its ambit-mates'.

But of course, two ambit-mates will rarely, if ever, have ambits whose areas  $a$  entirely overlap. More commonly, if we imagine two individuals A and B, their ambits will intersect incompletely (Fig. 2). Thus, although individual C resides in the intersection and is an ambit-mate of A and B, individual D only exists in the ambit of A, and individual E only resides in the ambit of B. However, because A and B are in the same group, and D is in A's group and E is in B's group, we must include D and E in the group as well. Thus, we can think of a group as the 'union' of all individuals whose ambits intersect with at least one other ambit (Fig. 2).

Such a definition is quite easy to implement algorithmically on a computer. We select a single individual from the field at random, and place him into 'group 1'. We then have the computer construct an 'ambit circle' of radius  $D$  around that individual, and place all individuals found inside the ambit circle into group 1 as well. We then eliminate the first individual from the pool, and select at random another individual in group 1. We construct an ambit around that individual and add to group 1 any ambit-mates



**Figure 2.** Schematic of how group membership is determined using the ambit method. All individuals have ambits, but for visual clarity, we show only the ambits of individual 'A' (dashed line) and individual 'B' (dotted line). Because 'A' and 'B' are group-mates due to overlapping ambits, all the other individuals depicted are also members of the same group. See text for further details.

of the second individual that were not yet included. This second individual is then eliminated from consideration and a third individual from group 1 is selected, and the process is continued. Eventually, all individuals in group 1 have been detected and eliminated from the field. We then randomly select a new individual and place him into 'group 2', and the process is repeated for all members of the population who were not part of group 1. We continue in this way until all individuals on the field have been assigned to one and only one of the groups.

### Simulations

To demonstrate how the 'ambit' method works, we generated five different 'test case' configurations (Fig. 3) for populations of  $I = 100$  individuals with fixed ambit radii of  $\hat{D} = 5$  BL. Thus, all individuals within a group meet our definition of being within the ambit of at least one other member of that group. We let  $\hat{G}$  represent the programmed group size (in numbers of individuals),  $\hat{\gamma}$  represent the programmed number of groups in the population and  $s$  represent the programmed separation among groups (in units of BL). The separation  $s$  of groups refers to the distance between group edges; thus, no individual was closer than  $s$  BL from someone in any other group. We tested our method in five cases: many small groups close together (Fig. 3a), many small groups far apart (Fig. 3c), a few large groups close together (Fig. 3b), a few large groups far apart (Fig. 3d), and one single large group (Fig. 3e). The values of the parameters used to generate the groups are shown in Table 1. For all groups, every individual's velocity vector was identical (groups were perfectly aligned), but the direction of travel for each group was randomly generated. We replicated each of these five population configurations 10 times. Thus, all reported statistics (e.g. group polarity) are sample means with  $N = 10$  replicates. These simulations allowed us to compare the results of our method when actual group membership is known a priori.

We tested the ambit method against these known group configurations in two ways. First, we provided the computer with four 'estimated' ambit radii  $D_{\text{est}}$  of varying sizes and compared group detection using these sizes. The estimated ambit radii we used were (1) the correct ambit ( $D_{\text{est}} = \hat{D}$ ), (2) an estimated ambit far too large ( $D_{\text{est}} = 3\hat{D}$ ), (3) an estimated ambit far too small ( $D_{\text{est}} = \hat{D}/2$ ), and (4) an estimated ambit that includes the whole population ( $D_{\text{est}} = \infty$ ). Thus, we were able to determine the effect of either including nongroup members in the calculations ( $D_{\text{est}} \gg \hat{D}$ ) or excluding some group members from the calculations ( $D_{\text{est}} \ll \hat{D}$ ). If the ambit method provides correct group detection, our results using the correct ambit ( $D_{\text{est}} = \hat{D}$ ) should result in the most accurate estimates of the simulated configurations.

The second test used only the two 'nearby group' cases, with groups separated by slightly more than one ambit radius from one another (8 BL from group edge to group edge, Fig. 3a, b). We chose these two conditions because they are the most challenging for our method due to close group proximity (see Figs 3, 4). Thus, we used the 'worst case' scenario for our sensitivity analysis; the closer the groups are to one another in space, the more difficult it is for the method to tell them apart. If the method succeeds here, it will succeed in the other cases. Consequently, we judged it unnecessary to repeat the procedure for the 16 BL condition. For the two 8 BL cases, then, we performed a sensitivity analysis to determine how results are affected by using estimated ambits  $D_{\text{est}}$  of varying sizes from 1 BL (much too small) to 25 BL (many times too large). Our purpose was to gain insight into how close  $D_{\text{est}}$  must be to  $\hat{D}$  to obtain accurate group detection. Given that we may not exactly know  $\hat{D}$  for living animals, it is critical to assess the method's robustness to observation error.

### Statistical Analysis

We computed three common aggregation indices (Parrish et al., 2002) for our simulated populations: number of groups, group expanse and group polarity. Number of groups is simply the number of unique groups identified by the computer algorithm using the estimated ambit radius  $D_{\text{est}}$ . Expanse is calculated for each group as the mean distance to the group's centre, and thus estimates group 'spread' (Huth & Wissel, 1992). Polarity is defined as mean angle deviation between the individual's velocity vector and that of the group's average velocity vector for each group (Huth & Wissel, 1992). We followed Viscido et al. (2007), Kolpas et al. (2013) and others in rescaling polarity so that  $0^\circ$  angle of deviation = 1 (perfect alignment) and  $180^\circ$  of angle deviation = 0 (complete misalignment).

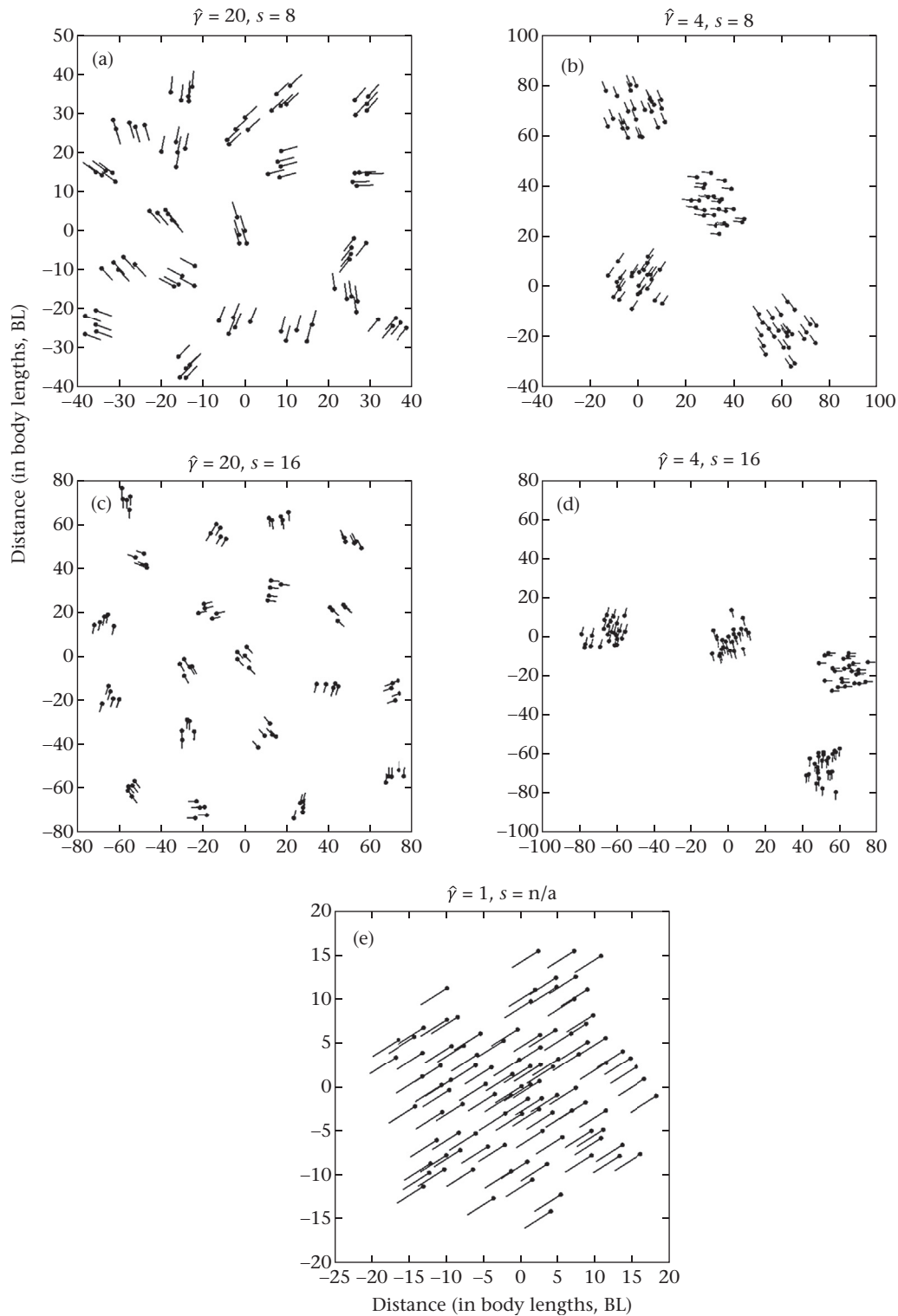
Both expanse and polarity were computed within a single simulated population for each group and then averaged across groups (number of groups can take on only a single value for each replicate). We computed the number of groups, mean expanse and mean polarity for each replicate and report the average across the 10 replicates for each index.

## RESULTS

How we assigned group membership had a strong impact on the number of groups detected (Fig. 4), the estimated group polarity (Fig. 5) and the estimated expanse of the group (Fig. 6). Regarding group detection, using an overly large ambit size ( $D_{\text{est}} = \{3\hat{D}, \infty\}$ ) caused the computer to underestimate the number of groups in all cases except when the entire population was seeded as part of a single group (Fig. 4). This effect was more pronounced when there were many small groups in the population (Fig. 4a, c) than when there were a few large groups in the population (Fig. 4b, d). Conversely, using an estimated ambit  $D_{\text{est}}$  that was too small overestimated the number of groups in all cases (Fig. 4), regardless of how many groups were seeded into the simulation. Using the correct estimated ambit size ( $D_{\text{est}} = \hat{D}$ ), however, resulted in a nearly perfect detection of groups every time (Fig. 4).

Because group statistics necessarily depend on which individuals are in the group, it is no surprise that the calculation of polarity and expanse was strongly affected by the estimated ambit size  $D_{\text{est}}$  used. For all configurations except one (when population = group), using  $D_{\text{est}} \gg \hat{D}$  caused an underestimate of group polarity (Fig. 5) and an overestimate of group expanse (Fig. 6). On the other hand, using  $D_{\text{est}} \ll \hat{D}$  provided incorrect estimates of expanse, but not of polarity (Figs 5, 6). Using  $D_{\text{est}} = \hat{D}$  always resulted in correct estimates of these group properties, regardless of group configuration (Figs 5, 6).

Our sensitivity analysis showed that using the ambit method can return nearly any result, depending on the ambit size  $D_{\text{est}}$  chosen, but that the ambit method is robust for a range of  $D_{\text{est}}$  values and is capable of returning the correct result throughout that range (Fig. 7). For example, with regard to number of groups detected, the ambit method returned the correct result (4 groups and 20 groups) across the range of  $D_{\text{est}} = [5, 10]$ , even though the populations were generated with a single ambit value ( $\hat{D} = 5$ ). Sensitivity results for the two group statistics estimated (polarity and expanse) showed similar patterns in three out of the four cases we examined, with plateaus occurring in the range of 5 to 10 BL (Fig. 7c, d, f). In the fourth case, polarity estimates for small groups, although there was no plateau, the 5–10 BL range still gave very similar (and nearly correct) results (Fig. 7e). Consequently, using estimated ambit sizes  $D_{\text{est}}$  on the interval  $[\hat{D}, 2\hat{D}]$  appears capable of providing the investigator with biologically meaningful group detection.



**Figure 3.** Sample population configurations for the five simulated conditions: (a) 20 groups of five individuals with at least 8 body lengths (BL) between group edges; (b) four groups of 25 individuals with at least 8 BL between group edges; (c) 20 groups of five individuals with at least 16 BL between group edges; (d) four groups of 25 individuals with at least 16 BL between group edges; (e) one group of 100 individuals. Groups were constructed with a known ambit size  $\hat{D} = 5$  BL in all cases. Dots indicate animal positions; lines indicate velocity vectors. All groups were perfectly aligned ( $\phi = 1$ ). Note: for visual clarity, axes are scaled differently in each case to accommodate the different group densities.

## DISCUSSION

This study clearly shows the importance of correctly assigning individuals to groups when calculating group properties and demonstrates that how we define groups can profoundly affect our

estimates of group properties. For example, when we lump individuals who are not actually group members together and call them a single aggregation, we will underestimate the polarity of the true group and overestimate its physical size (Figs 5, 6). Conversely, when we leave individuals out of the group who should

**Table 1**

Parameters used to set up the five programmed test population configurations for populations of  $I = 100$  individuals

Programmed group configuration	$\hat{G}$	$\hat{\gamma}$	$s$	Illustration
Many small groups, close together	5	20	8	Fig. 3a
Many small groups, far apart	5	20	16	Fig. 3c
Few large groups, close together	25	4	8	Fig. 3b
Few large groups, far apart	25	4	16	Fig. 3d
Whole population in one group	100	1	N/A	Fig. 3e

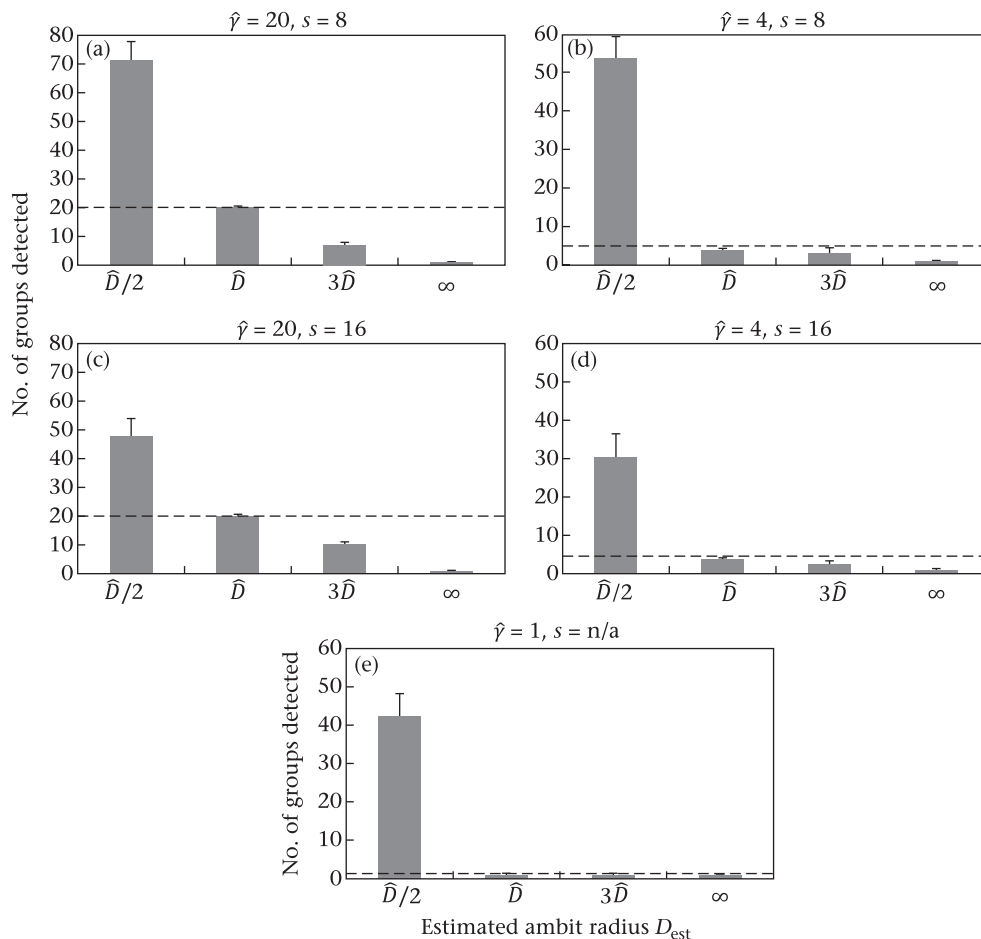
$\hat{G}$  represents the programmed number of individuals per group,  $\hat{\gamma}$  represents the programmed number of groups per 100-individual population and  $s$  represents the distance (in body lengths, BL) between group edges.

have been included, we will underestimate the group's physical size (Fig. 5). We note that, based on our simulated conditions, which included perfectly aligned groups, it was not possible to overestimate group polarity. For brevity, we did not program intermediate alignment values into our simulated groups, but leaving individuals out of the group can lead to an overestimate of polarity as well (data not shown). Consequently, investigators should employ the ambit method, or a method like it, to properly assign individuals to groups so that we can draw more informative biological conclusions about aggregation behaviour.

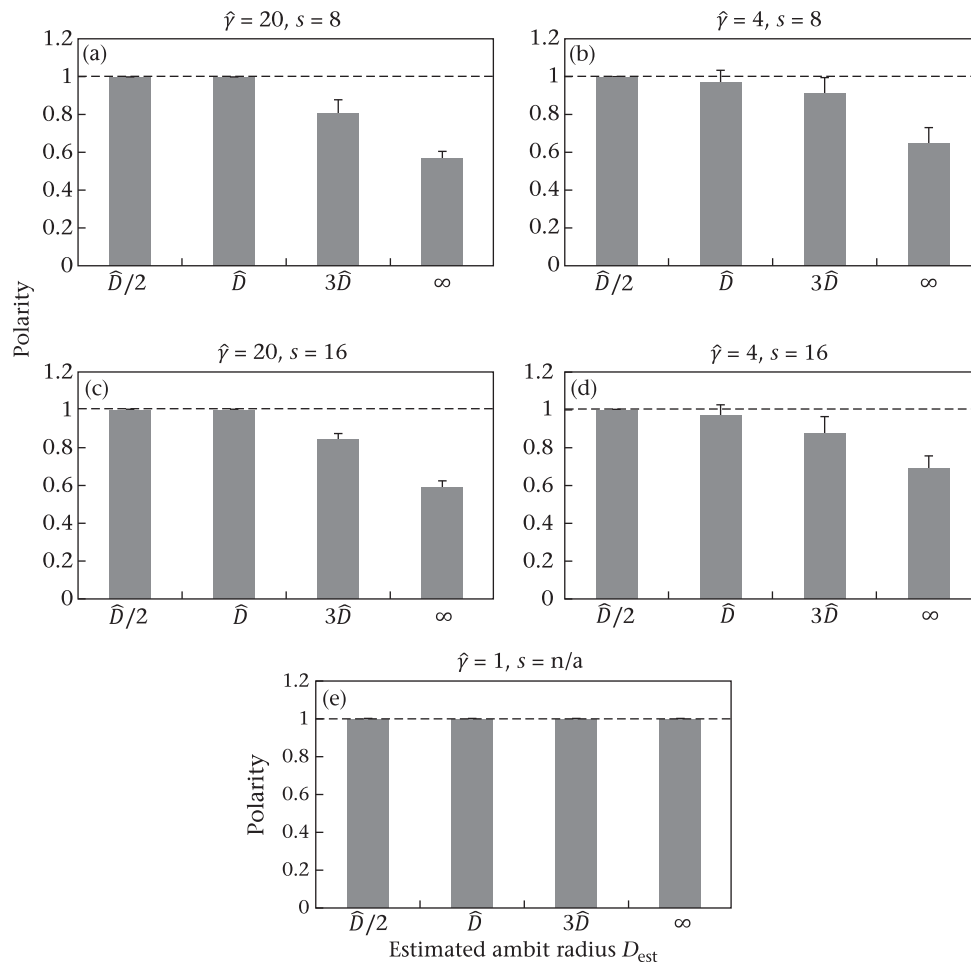
We began our study with two goals in mind: (1) developing a robust and biologically meaningful definition of an animal group

and (2) providing a computational algorithm that, with relatively minimal data, can correctly assign animals to groups in many cases. To accomplish the first goal, we defined an animal group as the set of all the members of a population who are in the interaction zone of at least one other member of the set during the observation period. To accomplish the second goal, we suggest using the 'ambit' as defined by Lloyd (1967) as a proxy for the interaction zone and aggregating individuals based on the union of their ambits. The method requires one to discover the interaction radius  $D$  of the animal in question, which is known for simulation models (as the zones of attraction, orientation and repulsion; e.g. Aoki, 1982; Couzin, Krause, James, Ruxton, & Franks, 2002; Huth & Wissel, 1992), but must be estimated using behavioural assays for living animals (e.g. Land & Layne, 1995a, b). However, once one has determined the ambit radius  $D$ , one need only discover the positional coordinates of animals to correctly assign group membership using our method.

Fortunately, although the interaction radius  $D$  of live animals may be difficult to estimate, our sensitivity analysis demonstrates that the ambit method is relatively robust across a range of values and does not require the exact ambit to be known (Fig. 7). Indeed, in our example simulations, the computer returned correct values for the number of groups, polarity and expanse for a range of ambit values from  $D$  to  $2D$ . This leads us to two tentative conclusions that require further analysis. First, the method will return



**Figure 4.** Mean number of groups detected using different values for estimated ambit radius  $D_{est}$  for (a) 20 groups of five individuals with at least 8 body lengths (BL) between group edges; (b) four groups of 25 individuals with at least 8 BL between group edges; (c) 20 groups of five individuals with at least 16 BL between group edges; (d) four groups of 25 individuals with at least 16 BL between group edges; (e) one group of 100 individuals. Error bars indicate +1 SD. The dashed line indicates the correct answer (i.e. what the simulation was programmed to produce). Note: for visual clarity, axes are scaled differently in each case to accommodate the different group densities.



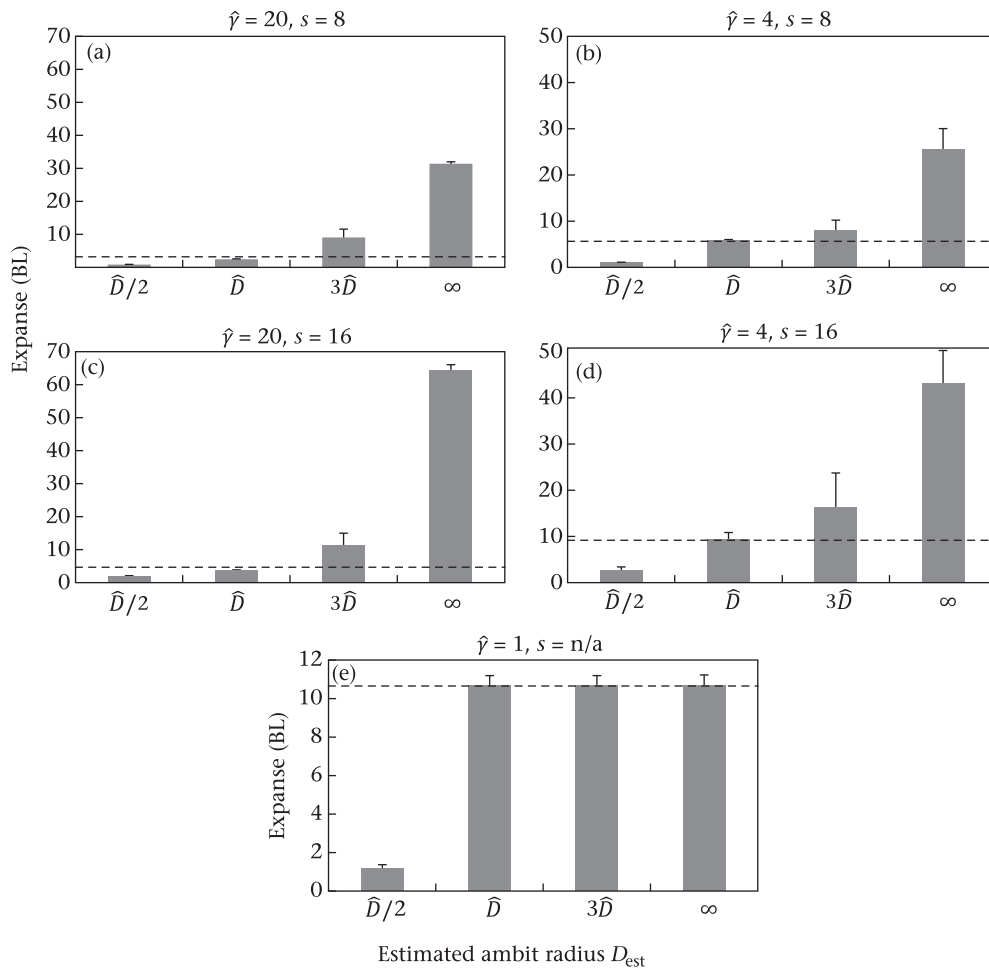
**Figure 5.** Mean polarity  $\phi$  of groups detected using different values for estimated ambit radius  $D_{\text{est}}$  for (a) 20 groups of five individuals with at least 8 body lengths (BL) between group edges; (b) four groups of 25 individuals with at least 8 BL between group edges; (c) 20 groups of five individuals with at least 16 BL between group edges; (d) four groups of 25 individuals with at least 16 BL between group edges; (e) one group of 100 individuals. Error bars indicate  $+1$  SD. The dashed line indicates the correct answer (i.e. what the simulation was programmed to produce).

biologically meaningful results over an ambit range that varies by as much as 100%, and second, it is better to slightly overestimate  $D$  than to underestimate it. However, we warn the reader that we could not comprehensively test all possible group configurations, and there will certainly be some conditions, such as groups with elongated shapes, where overestimates of  $D$  will cause more difficulties. Thus, although the method appears to be robust and certainly works well for simulations, we urge those studying animals in the field to obtain the most accurate estimate of ambit radius possible.

Although the ‘ambit method’ we present here appears to be robust and capable of giving good results in the generic cases outlined above, we readily acknowledge that there may be a variety of population configurations for which the method will not work. For example, [Hristov, Betke, and Kunz \(2008\)](#) and [Reichard et al. \(2009\)](#) described long streams of free-tailed bats, *Tadarida brasiliensis*, emerging from a cave on foraging bouts. These streams can be hundreds of metres long and take on highly distinctive shapes (N. Hristov, personal communication). Using the ambit method, one would conclude that the entire population is composed of a single group. And yet, there is likely no connection, even indirectly, among individuals at opposite ends of the elongated swarm. Similarly, solitary individuals who are located in between two clearly distinct, noninteracting groups, but who are barely within the ambit of each

group, would cause the method to call the whole set of individuals ‘one group’, when most biologists would identify them as two distinct groups and a straggler. Selecting a smaller estimated ambit size could solve this problem, but might cause the two groups to be identified as many small groups, which would introduce other inaccuracies. Clearly, the ambit method is one tool in the arsenal, but other methods will be required to characterize unique configurations such as those described above.

The ambit method is useful because it is both easy to implement and frequently provides valuable insights; however, the method’s simplicity imposes several limits on the types of groups that can be distinguished. For example, the method does not take heading or speed into account, because it assumes allelomimesis, which usually causes neighbouring animals to follow similar trajectories. Thus, the ambit method cannot distinguish two groups that are near each other but going in different directions, nor can it differentiate between a school and a shoal ([Shaw, 1978](#)). Similarly, the ambit method cannot tell whether, during a single time step, a moving group is passing by a stationary nongroup member. In such situations, additional methods would be required, such as computing the cross-correlation of individual versus group heading ([Kolpas et al., 2013](#)). Whether these more involved methods are required, or whether the ambit method is sufficient, will depend on the biological conditions of the study.



**Figure 6.** Mean expanse of groups detected using different values for estimated ambit radius  $D_{est}$  for (a) 20 groups of five individuals with at least 8 body lengths (BL) between group edges; (b) four groups of 25 individuals with at least 8 BL between group edges; (c) 20 groups of five individuals with at least 16 BL between group edges; (d) four groups of 25 individuals with at least 16 BL between group edges; (e) one group of 100 individuals. Error bars indicate +1 SD. The dashed line indicates the correct answer (i.e. what the simulation was programmed to produce). Note: for visual clarity, axes are scaled differently in each case to accommodate the different group densities.

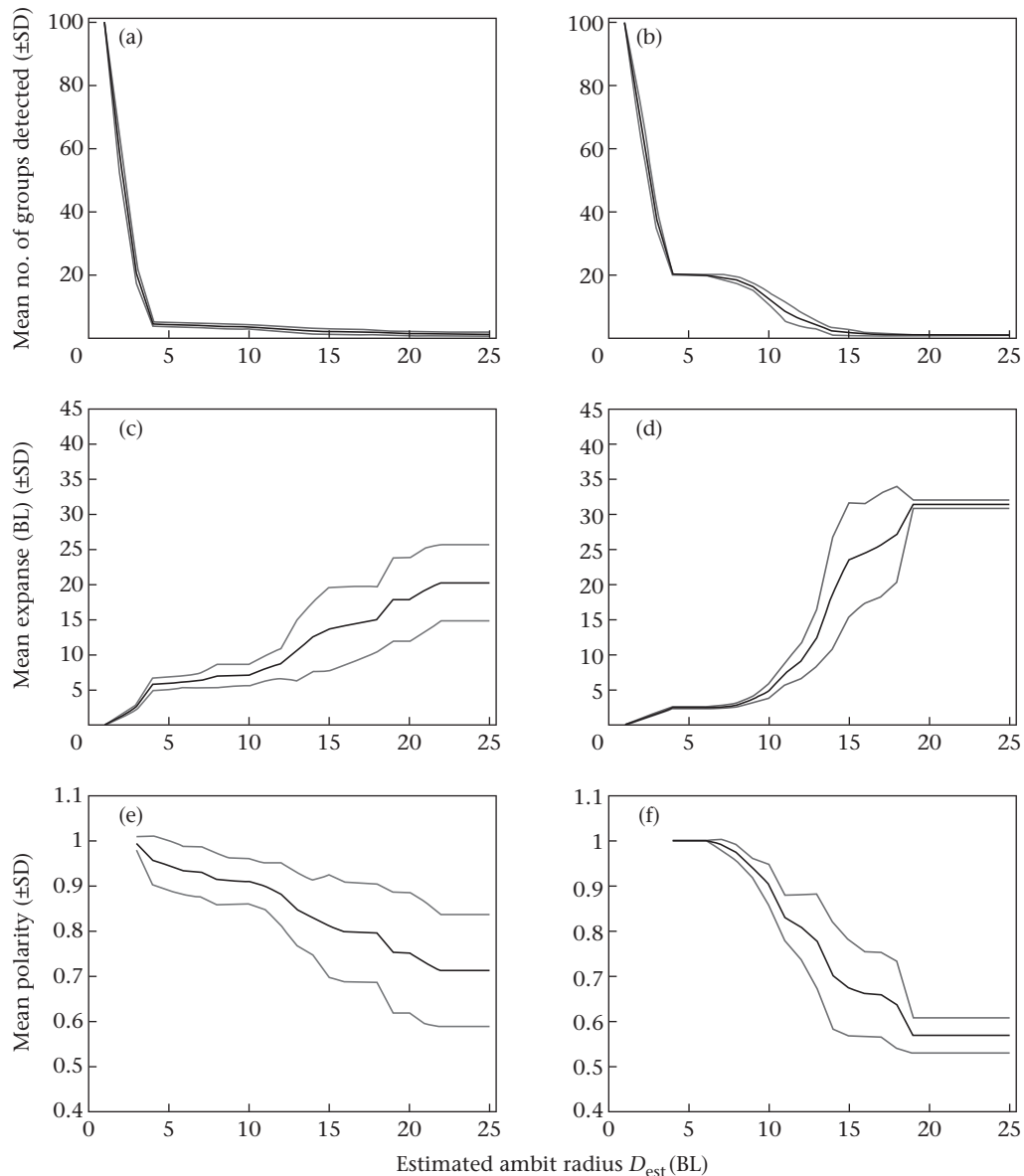
The ambit method uses an algorithm for group detection based on algorithms in some earlier studies. For example, although each study employed slightly different terminology, Gueron et al. (1996) used the same method in simulations of ungulate herds, as did Flierl et al. (1999) with invertebrate swarm simulations, and Gautrais et al. (2008) for fish-schooling simulations. More recently, Romey and Vidal (2013) used a very similar method (agglomerative clustering using nearest-neighbour distances) to identify group membership in simulations of animal swarms, based on techniques from a much earlier paper (Ward, 1963). This ‘greedy’ method has been known for many years in the computational literature, but so far has remained relatively unknown to biologists, demonstrating the importance of interdisciplinary collaboration when studying group behaviour. The key feature in all these cases is the use of positional information and interaction distances to infer group membership, which, as we have shown, is necessary if one is to draw relevant biological conclusions about animal groups (Figs 5–7). Therefore, we present the method not as a new technique, but as an example of how important proper identification of group membership is, and as one possible, robust algorithm for group detection.

Animal aggregation is one of the most studied aspects of animal behaviour, with a long history of both live-animal and simulation studies. In many cases, investigators have reported group-based statistics such as polarity (Kolpas et al., 2013; Reuter

& Breckling, 1994), group speed (Couzin et al., 2002) and expanse (Huth & Wissel, 1992; Viscido et al., 2007). However, because ‘the group’ had not been formally defined, one cannot be certain that these variables were measured in ways that permit easy comparison across studies. Indeed, as Parrish et al. (2002) pointed out, all individuals in the field (i.e. the population) are often counted as being ‘in the group’ (e.g. Viscido, Miller, & Wetthey, 2002). Although in some cases this may be justified (e.g. Couzin et al., 2002), frequently calling all visible animals ‘group members’ has questionable biological validity. And because a common definition does not exist, even when group detection methods are employed, the algorithms used to do so may be widely different among studies. For example, Gueron et al. (1996) and Romey and Vidal (2013) used an algorithm much like the one we propose in this paper, but other investigators have used very different algorithms (e.g. Heard, 1992; Stöcker, 1999). Clearly, animal aggregation studies would benefit from greater standardization.

The formal definition of ‘a group’ and an algorithm for detecting groups would merely be an argument in semantics were it not for the most critical finding we present: that for a given population of animals arranged in a single aggregation configuration, when we change how we define the group, we can completely change the biological conclusions we draw (Figs 4–6). Even for the method we suggest, greatly over- or underestimating





**Figure 7.** Mean number of detected groups, polarity and expanse for populations with four groups of 25 individuals each (a, c, e) and 20 groups of five individuals each (b, d, e) across a range of estimated ambit sizes from  $D_{\text{est}} = 1$  to 25 body lengths (BL) ( $D_{\text{est}} = \bar{D}/5$  to  $5\bar{D}$ ). Black lines indicate means; grey lines indicate  $\pm 1$  SD.

the ambit radius can lead one to opposite conclusions. For example, by changing the estimated ambit radius for a population containing 20 groups of five individuals, we can cause the computer to detect from 1 to 100 groups, with expanses ranging from 0 to 45 BL, and polarities ranging from 0.5 to 1. Note that in this case, and the others we present, the positions of the animals did not change. And yet, depending on how we defined the group, we could draw such clearly incompatible conclusions as small, highly aligned groups or large, disorganized groups. Thus, the definition of the group and the algorithm used to detect the group are clearly critical not just for semantic reasons, but because they affect the conclusions we draw about the consequences of animal behaviour.

Animal group formation is a complex phenomenon that has captured the imagination of both scientists and lay people for decades (Couzin, 2007; Couzin & Krause, 2003; Parrish & Edelstein-Keshet, 1999). Understandably, this has led to a wide array of studies both in the field (e.g. Ballerini et al., 2008; Heard, 1992; Hristov, Allen, & Chadwell, 2013; Hristov, Betke, Theriault, Bagchi,

& Kunz, 2010) and in simulation (e.g. Flierl et al., 1999; Li, Kolpas, Petzold, & Moehlis, 2009; Reynolds, 1987; Romey & Vidal, 2013). However, the lack of a common definition of ‘a group’, as well as the absence of a standard method for group detection, has led to a confusing array of studies whose findings are difficult to compare. Although we propose one workable method here, we do not argue that all investigators should use it in every case, because no method can return meaningful results in all situations. Rather, we encourage investigators to formally define their meaning when they use terms such as ‘group’ or ‘aggregation’, and to present a clear algorithmic method for group detection. Explicitly defining the group will increase the rigour of animal aggregation research and greatly enhance our ability to make connections across studies.

#### Acknowledgments

We thank the undergraduates at Winston-Salem State University who assisted in the development of the code used in this paper, including A. Kennedy, S. Lea and W. Hill. We thank N. Hirstov for

helpful discussions of theory and for commenting on a previous draft. This paper also benefited from the keen insights of two anonymous referees. Funding for this project was provided by National Science Foundation grant IOS-1149302 to S. Viscido.

## References

- Aoki, I. (1982). A simulation study on the schooling mechanism in fish. *Bulletin of the Japanese Society of Scientific Fisheries*, 48, 1081–1088.
- Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., et al. (2008). Interaction ruling animal collective behavior depends on topological rather than metric distance: evidence from a field study. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 1232–1237.
- Barbour, M. G., Burk, J. H., & Pitts, W. D. (1987). *Terrestrial plant ecology* (2nd ed.). Menlo Park, CA: Benjamin Cummings.
- Beauchamp, G. (2012). Flock size and density influence speed of escape waves in semipalmated sandpipers. *Animal Behaviour*, 83, 1125–1129.
- Beauchamp, G. (2014). *Social predation: How group living benefits predators and prey* (1st ed.). London, U.K.: Academic Press.
- Burger, J. (1981). A model for the evolution of mixed-species colonies of Ciconiiformes. *Quarterly Review of Biology*, 56, 143–167.
- Cambui, D. S., & Rosas, A. (2012). Density induced transition in a school of fish. *Physica A: Statistical Mechanics and Its Applications*, 391, 3908–3914.
- Couzin, I. (2007). Collective minds. *Nature*, 445, 715.
- Couzin, I. D., & Krause, J. (2003). Self-organization and collective behavior in invertebrates. *Advances in the Study of Behavior*, 32, 1–75.
- Couzin, I. D., Krause, J., James, R., Ruxton, G. D., & Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, 218, 1–11.
- Flierl, G., Grünbaum, D., Levin, S., & Olson, D. (1999). From individuals to aggregations: the interplay between behavior and physics. *Journal of Theoretical Biology*, 196, 397–454.
- Gautrais, J., Jost, C., & Theraulaz, G. (2008). Key behavioural factors in a self-organised fish school model. *Annales Zoologici Fennici*, 45, 415–428.
- Götmark, F., Winkler, D. W., & Andersson, M. (1986). Flock-feeding on fish schools increases individual success in gulls. *Nature*, 319, 589–591.
- Gueron, S., Levin, S. A., & Rubenstein, D. I. (1996). The dynamics of herds: from individuals to aggregations. *Journal of Theoretical Biology*, 182, 85–98.
- Hamilton, W. D. (1971). Geometry for the selfish herd. *Journal of Theoretical Biology*, 31, 295–311.
- Heard, D. C. (1992). The effect of wolf predation and snow cover on musk-ox group size. *American Naturalist*, 139, 190–204.
- Hristov, N. I., Allen, L. C., & Chadwell, B. A. (2013). New advances in the study of group behavior in bats. In R. A. Adams, & S. C. Pedersen (Eds.), *Bat evolution, ecology and conservation* (pp. 271–291). New York, NY: Springer Science Press.
- Hristov, N. I., Betke, M., & Kunz, T. H. (2008). Applications of thermal infrared imaging for research in aerocology. *Integrative and Comparative Biology*, 48, 50–59.
- Hristov, N. I., Betke, M., Theriault, D. E. H., Bagchi, A., & Kunz, T. H. (2010). Seasonal variation in colony size of Brazilian free-tailed bats at Carlsbad Cavern based on thermal imaging. *Journal of Mammalogy*, 91, 183–192.
- Huth, A., & Wissel, C. (1992). The simulation of the movement of fish schools. *Journal of Theoretical Biology*, 156, 365–385.
- Jarman, P. J. (1987). Group size and activity in eastern grey kangaroos. *Animal Behaviour*, 35, 1044–1050.
- Katz, Y., Tunström, K., Ioannou, C. C., Huepe, C., & Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 18720–18725.
- Kolpas, A., Busch, M., Li, H., Couzin, I. D., Petzold, L., & Moehlis, J. (2013). How the spatial position of individuals affects their influence on swarms: a numerical comparison of two popular swarm dynamics models. *PLoS One*, 8, e58525.
- Land, M., & Layne, J. (1995a). The visual control of behaviour in fiddler crabs. I. Resolution, thresholds and the role of the horizon. *Journal of Comparative Physiology A*, 177, 81–90.
- Land, M., & Layne, J. (1995b). The visual control of behaviour in fiddler crabs. II. Tracking control systems in courtship and defence. *Journal of Comparative Physiology A*, 177, 91–103.
- Li, H., Kolpas, A., Petzold, L., & Moehlis, J. (2009). Parallel simulation for a fish schooling model on a general-purpose graphics processing unit. *Concurrency and Computation: Practice and Experience*, 21, 725–737.
- Lloyd, M. (1967). Mean crowding. *Journal of Animal Ecology*, 36, 1–30.
- Parrish, J. K. (1991). Do predators shape fish schools: interactions between predators and their schooling prey. *Netherlands Journal of Zoology*, 42, 358–370.
- Parrish, J. K., & Edelstein-Keshet, L. (1999). Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science*, 284, 99–101.
- Parrish, J. K., Viscido, S. V., & Grünbaum, D. (2002). Self-organized fish schools: an examination of emergent properties. *Biological Bulletin*, 202, 296–305.
- Petit, D. R., & Bildstein, K. L. (1987). Effect of group size and location within the group on the foraging behavior of white ibises. *Condor*, 89, 602–607.
- Pitcher, T. J., & Parrish, J. K. (1993). Functions of shoaling behaviour in teleosts. In T. Pitcher (Ed.), *Behaviour of teleost fishes* (Vol. 2, pp. 363–439). London, U.K.: Chapman & Hall.
- Reichard, J. D., Gonzalez, L. E., Casey, C. M., Allen, L. C., Hristov, N. I., & Kunz, T. H. (2009). Evening emergence behavior and seasonal dynamics in large colonies of Brazilian free-tailed bats. *Journal of Mammalogy*, 90, 1478–1486.
- Reuter, H., & Breckling, B. (1994). Self-organization of fish schools: an object-oriented model. *Ecological Modelling*, 75, 147–159.
- Reynolds, C. W. (1987). Flocks, herds, and schools: a distributed behavioral model. *SIGGRAPH Computer Graphics*, 21, 25–34.
- Ricklefs, R., & Relyea, R. (2014). *Ecology: The economy of nature* (7th ed.). New York, NY: W. H. Freeman.
- Ritz, D. A. (1994). Social aggregation in pelagic invertebrates. *Advances in Marine Biology*, 30, 156–216.
- Romey, W. L., & Vidal, J. M. (2013). Sum of heterogeneous blind zones predict movements of simulated groups. *Ecological Modelling*, 258, 9–15.
- Salierno, J., Gipson, G., & Kane, A. (2008). Quantitative movement analysis of social behavior in mummichog, *Fundulus heteroclitus*. *Journal of Ethology*, 26, 35–42.
- Shaw, E. (1978). Schooling fishes. *American Scientist*, 66, 166–175.
- Stöcker, S. (1999). Models for tuna school formation. *Mathematical Biosciences*, 156, 167–190.
- Vabø, R., & Nøttestad, L. (1997). An individual-based model of fish school reactions: predicting antipredator behaviour as observed in nature. *Fisheries Oceanography*, 6, 155–171.
- Viscido, S. V., Miller, M., & Wethey, D. S. (2002). The dilemma of the selfish herd: the search for a realistic movement rule. *Journal of Theoretical Biology*, 217, 183–194.
- Viscido, S. V., Parrish, J. K., & Grünbaum, D. (2007). Factors influencing the structure and maintenance of fish schools. *Ecological Modelling*, 206, 153–165.
- Viscido, S. V., & Wethey, D. S. (2002). Quantitative analysis of fiddler crab flock movement: evidence for 'selfish herd' behaviour. *Animal Behaviour*, 63, 735–741.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236–244.
- Williams, G. C. (1964). Measurement of consociation among fishes and comments on the evolution of schooling. *Publications of the Museum, Michigan State University Biological Series*, 2, 351–383.
- Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Cambridge, MA: Harvard University Press.
- Wrona, F. J., & Dixon, R. W. J. (1991). Group size and predation risk: a field analysis of encounter and dilution effects. *American Naturalist*, 137, 186–201.
- Zheng, M., Kashimori, Y., Hoshino, O., Fujita, K., & Kambara, T. (2005). Behavior pattern (innate action) of individuals in fish schools generating efficient collective evasion from predation. *Journal of Theoretical Biology*, 235, 153–167.