Animal Behaviour 82 (2011) 911–921

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Commentary

Assessing dominance hierarchies: validation and advantages of progressive evaluation with Elo-rating

Christof Neumann [a,b,c,d,*], Julie Duboscq [a,b], Constance Dubuc [a,b,c], Andri Ginting [a,e], Ade Maulana Irwan [a,e], Muhammad Agil [e], Anja Widdig [c,d,1], Antje Engelhardt [a,b,1]

[a] Junior Research Group of Primate Sexual Selection, Department of Reproductive Biology, German Primate Center
[b] Courant Research Center 'Evolution of Social Behavior', Georg-August-Universität, Germany
[c] Junior Research Group of Primate Kin Selection, Department of Primatology, Max-Planck-Institute for Evolutionary Anthropology, Germany
[d] Institute of Biology, Faculty of Bioscience, Pharmacy and Psychology, University of Leipzig, Germany
[e] Faculty of Veterinary Medicine, Bogor Agricultural University, Indonesia

ARTICLE INFO

Article history:
Received 4 March 2011
Initial acceptance 21 April 2011
Final acceptance 29 June 2011
Available online 12 August 2011
MS. number: 11-00183

Keywords:
David's score
dominance hierarchy
dominance rank
Elo-rating
hierarchy stability
I&SI
Macaca mulatta
Macaca nigra
methodology

Dominance is one of the most important concepts in the study of animal social behaviour. Dominance hierarchies in groups arise from dyadic relationships between dominant and subordinate individuals present in a social group (Drews 1993). High hierarchical rank or social status is often associated with fitness benefits for individuals (e.g. Côté & Festa-Bianchet 2001; von Holst et al. 2002; Widdig et al. 2004; Engelhardt et al. 2006), and hierarchies can be found in most animal taxa including insects (e.g. Kolmer & Heinze 2000), birds (e.g. Kurvers et al. 2009) and mammals (e.g. Keiper & Receveur 1992).

The analysis of dominance has a long-standing history (Schjelderup-Ebbe 1922; Landau 1951), and a great number of methods to assess hierarchies in animal societies are currently available (reviewed in de Vries 1998; Bayly et al. 2006; Whitehead 2008). Although differing in calculation complexity, all ranking methods presently used in studies of behavioural ecology are based on interaction matrices. For this, a specific type of behaviour or interaction, from which the dominance/subordinance relationship of a given dyad can be deduced, is tabulated across all individuals (see for example, Vervaecke et al. 2007). This matrix can either be reorganized as a whole in order to optimize a numerical criterion (e.g. I&SI: de Vries 1998; minimizing entries below the matrix diagonal: Martin & Bateson 1993), or alternatively, an individual measure of success calculated for each animal present (e.g. David's score: David 1987; CBI: Clutton-Brock et al. 1979). In the latter case, a ranking can be generated by ordering the individual scores obtained.

Although calculations of dominance hierarchies are routinely undertaken in many studies of behavioural ecology, and although there have been numerous methodological developments in this area (e.g. Clutton-Brock et al. 1979; David 1987; de Vries 1998), there are still a number of obstacles and limitations scientists have to tackle when analysing dominance relationships. This is mainly

* Correspondence: C. Neumann, Junior Research Group of Primate Sexual Selection, German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany.
E-mail address: cneumann@dpz.eu (C. Neumann).
[1] Equally contributed as senior authors.

doi:10.1016/j.anbehav.2011.07.016

because the methods commonly used can often not be applied to highly dynamic animal societies, or to sparse data sets, and because methods based on interaction matrices need to fulfil certain criteria to generate reliable results. Generally, many researchers may not be aware of some of the problems that are associated with the application of such methods to their data sets, which may in the worst case lead to the misinterpretation of results.

An alternative method that can overcome the shortcomings of matrix-based methods is Elo-rating. Developed by and named after Arpad Elo (Elo 1978), it is used for ratings in chess and other sports (e.g. Hvattum & Arntzen 2010), but has rarely been used in behavioural ecology (but see Rusu & Krackow 2004; Pörschmann et al. 2010). The major difference to commonly used ranking methods is that Elo-rating is based on the sequence in which interactions occur, and continuously updates ratings by looking at interactions sequentially. As a consequence, there is no need to build up complete interaction matrices and to restrict analysis to defined time periods. Ratings (after a given start-up time) can be obtained at any point in time, thus allowing monitoring of dominance ranks on the desired timescale.

The major aim of this paper is to promote Elo-rating among behavioural ecologists by illustrating its advantages over common methods, and by validating its reliability for assessing dominance rank orders, particularly in highly dynamic social systems. By providing the necessary computational tools along with an example (see Supplementary Material), we also make Elo-rating user friendly. In the following, we start with an introduction to the procedures of Elo-rating. We then show that with Elo-rating it is easy to track changes in social hierarchies, which may be overlooked with matrix-based methods, and point out several general advantages of Elo-rating over matrix-based methods. To demonstrate the benefits of Elo-rating empirically, we present the results of a reanalysis of one of our own previously published data sets. Finally, we validate the reliability and robustness of Elo-rating by comparing the performance of this method with those of two currently widely used ranking methods, the I&SI method and the David's score, using empirical data and reduced data sets that mimic sparse data.

ELO-RATING PROCEDURE

Elo-rating, in contrast to commonly used methods, is not based on an interaction matrix, but on the sequence in which interactions occur. At the beginning of the rating process, each individual starts with a predefined rating, for example a value of 1000. The amount chosen here has no effect on the differences in ratings later: the relative distances between individual ratings will remain identical (Albers & de Vries 2001). After each interaction, the ratings of the two participants are updated according to the outcome of the interaction: the winner gains points and the loser loses points. The number of points gained and lost during one interaction depends on the expectation of the outcome (i.e. the probability that the higher-rated individual wins, Elo 1978) prior to this interaction. Expected outcomes lead to smaller changes in ratings than unexpected outcomes (Fig. 1). Depending on whether the higher-rated individual wins or loses an interaction, ratings are updated according to the following formulae.

Higher-rated individual wins:

$$\text{WinnerRating}_{\text{new}} = \text{WinnerRating}_{\text{old}} + (1-p)\times k \quad (1)$$

$$\text{LoserRating}_{\text{new}} = \text{LoserRating}_{\text{old}} - (1-p)\times k \quad (2)$$

Lower-rated individual wins (against the expectation):

$$\text{WinnerRating}_{\text{new}} = \text{WinnerRating}_{\text{old}} + p\times k \quad (3)$$

$$\text{LoserRating}_{\text{new}} = \text{LoserRating}_{\text{old}} - p\times k \quad (4)$$

where p is the expectation of winning for the higher-rated individual, which is a function of the absolute difference in the ratings of the two interaction partners before the interaction (Fig. 1; see also Elo 1978; Albers & de Vries 2001). k is a constant and determines the number of rating points that an individual gains or loses after a single encounter. Its value is usually set between 16 and 200 and, once chosen, remains at this value throughout the rating

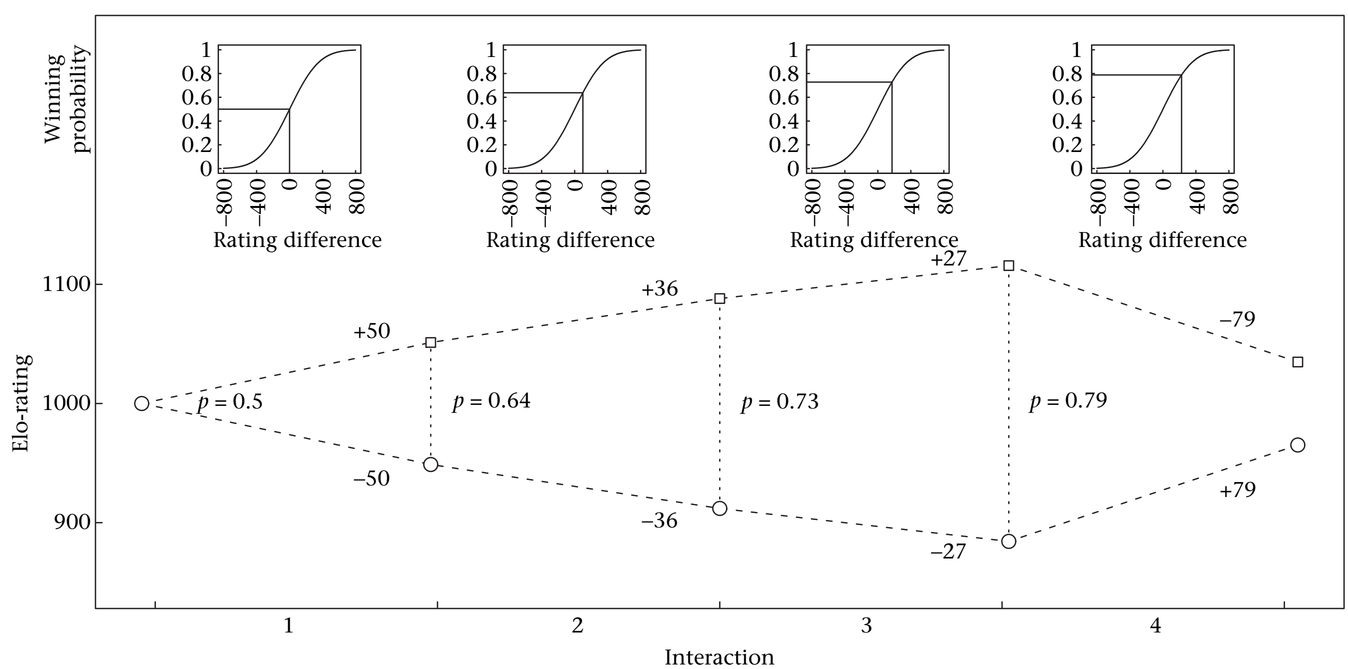

Figure 1. Graphical illustration of Elo-rating principles. Two individuals A (squares) and B (circles) interact four times of which the first three interactions are won by A and the fourth is won by B. The number of points gained/lost depends on the probability that the higher-rated individual wins the interaction (see text for details). The winning probability (p) is a function of the difference in Elo-ratings before the interaction (dotted vertical lines). As the difference in ratings increases with each interaction so does the chance of A winning. A graphical way to obtain the winning chance is depicted in the inset figures. A detailed description of this example can be found in Appendix 1.