



# Observer gender and observation bias in animal behaviour research: experimental tests with red-backed salamanders

DAVID M. MARSH & TERESA J. HANLON

Department of Biology, Washington and Lee University

(Received 25 August 2003; initial acceptance 5 December 2003;  
final acceptance 9 February 2004; published online 27 October 2004; MS. number: A9684)

Most studies of animal behaviour are based on direct observations of behaviour in a natural or laboratory context. While the potential for observation biases has often been discussed, there have been few quantitative analyses of the kinds of biases that may affect behavioural data. We used multiple observers of aggression and foraging behaviour in red-backed salamanders, *Plethodon cinereus*, to investigate observation biases related to the gender of the observer. We divided observers so that half were aware of the sex of the salamanders and the other half were kept blind to salamander sex. We then used comparisons between blind and unblind treatments to determine the magnitude of observation biases. We found little evidence for bias due to differential perception of male and female animals by men and women observers ('gender identification bias') for any of the behaviours examined. Although the expectations of men and women about salamander behaviour did sometimes differ, we also found no evidence that observers' expectations affected their observations of salamander behaviour. However, for one component of aggressive behaviour and one component of foraging behaviour, men observed higher overall frequencies of behaviour than did women, regardless of the sex of the salamander. Additionally, for three components of aggressive behaviour, both men and women recorded greater frequencies of behaviour when they were aware of the sex of the salamander than when they were unaware of the salamanders' sex. For two components of foraging behaviour, there were no significant biases with respect to the observers' gender or their knowledge of the sex of the salamander. These results suggest that observation biases with respect to observer gender may exist for some behavioural variables, and that these biases may be relatively subtle. Although conducting blind experiments may eliminate some types of bias, our results suggest that behaviours should also be carefully screened for consistency and repeatability prior to formal data collection, even for blind experiments.

© 2004 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

The scientific study of animal behaviour assumes that observations about behaviour reflect what animals are actually doing. It also assumes that these observations can be replicated by different observers. Unfortunately, making unbiased observations of animal behaviour is difficult for a variety of reasons. One problem is that the behaviour of many animals can be affected by the presence of the observer (Rosenthal 1966; Ollason & Dunnet 1980; Caine 1990). For this reason, many observations require long acclimation periods before reliable data can be collected (van Lawick-Goodall 1968; Rasmussen 1979; Jordan & Burghardt 1986). In addition, observers may have a tendency to assign human characteristics to nonhuman animals (i.e. anthropomorphise). The science of animal

behaviour has a long history of discussions about the perils of anthropomorphism (Morgan 1890; Thorndike 1898; Washburn 1908), and several important movements in the science have aimed to reduce its influence on the observation process by relying on quantitative data (Watson 1924; Barlow 1968; Altmann 1974).

Recently, a number of biologists have raised concerns about the potential for the gender of the scientist to influence the way that animal behaviour is studied (Gowaty 1997; Zuk 2002). Specifically, several authors have suggested that men and women may tend to focus on different aspects of behaviour, and may interpret the same behavioural data in very different ways. As a result, the historical underrepresentation of women in science may have biased how we think about animal behaviour. For example, Holmes & Hitchcock (1997) showed that women are much more likely than men to study mammals, whereas men are more likely to study fish,

Correspondence: D. M. Marsh, Department of Biology, Washington and Lee University, Lexington, VA 24450, U.S.A. (email: [marshd@wlu.edu](mailto:marshd@wlu.edu)).

amphibians and insects. In addition, for primates (although not for other taxa), women tend to focus on female animals and men tend to focus on male animals (Holmes & Hitchcock 1997). Other authors have presented anecdotal accounts of how behavioural data may have been interpreted in different ways by male and female scientists (Lawton et al. 1997; Pierotti et al. 1997), or how some ideas about animal behaviour may have been ignored because they went against prevailing ideas about human gender (Hrdy 1981; Zuk 2002).

The focus of these accounts has been on gender-related biases in the choice of what to study and in how data are interpreted. Few have questioned whether the gender of the scientist can affect the data collected in any given study. There are several ways in which observer gender could potentially influence data collection. First, men and women may view male and female animals differently. This 'gender identification bias' could lead to men and women reaching different conclusions about the ways in which male and female animals differ. Second, men and women may have overall differences in how they observe behaviours such as aggression, foraging, or courtship. This kind of bias could be independent of the sex of the animals being observed. Third, men and women may have different expectations about the behaviour of male and female animals. Such differing expectations could lead to differences in the behavioural data collected.

These kinds of gender-related biases may seem unlikely, and certainly the propensity for such biases will depend on the subjectivity of the behaviours being observed. However, rigorous testing across observers is the only way to determine whether or not behavioural observations are objective and repeatable. Differences in data collection between men and women have been found in some other fields. Psychologists have found that men and women may differ in their assignment of subjective 'hostility scores' for the same subjects (Davidson et al. 1996). In medicine, men and women have been shown to differ in their subjective ratings of the pain levels of others (Robinson & Wise 2003), and in their diagnosis of more objective conditions such as laryngeal papilloma (Behar & Todd 1999).

We investigated whether similar kinds of gender-related observation bias could occur in animal behaviour research. As an experimental system, we used 49 male and 58 female university students making laboratory observations on the aggression and foraging behaviour of red-backed salamanders, *Plethodon cinereus*. Half of these observers were made aware of the sex of the salamanders being observed, and the other half were unaware of the sex of the salamanders. We then tested for the presence of four kinds of gender-related observation biases. We asked (1) whether observers showed gender identification bias, defined here as differential observation of the behaviours of male and female animals by men and women; (2) whether men and women observers scored behaviours differently independent of the sex of the salamanders being observed; (3) whether simply knowing the sex of the salamander affected the frequency of recorded behaviours; and (4) how the sex of the observer was related to his or her a priori expectations about how salamanders

would behave and whether these expectations influenced the data recorded.

## METHODS

### Study System

We used red-backed salamanders for this experiment because they can be collected in large numbers and their behaviour can be easily observed in the laboratory. Red-backed salamanders are a terrestrial, lungless, plethodontid salamander found in woodlands throughout the eastern United States and Canada. Their social behaviour has been well studied in both the laboratory and the field (see Jaeger & Forester 1993 for review). These studies have determined that both male and female red-backed salamanders defend territories during at least part of the year (Mathis 1990). Aggressive behaviour is also displayed by both sexes, and territorial residents may bite intruders or engage in aggressive displays (Jaeger 1984). Foraging behaviour in red-backed salamanders has also been detailed in previous studies (Jaeger et al. 1981, 1982). Red-backed salamanders appear to forage primarily by sight, although they can also use smell to locate nonmotile prey (David & Jaeger 1981).

We collected sexually mature red-backed salamanders from the Jefferson National Forest in Giles County, Virginia, U.S.A., in September and October 2002. Salamanders from this population have been used in much of the previous behavioural work on red-backed salamanders (e.g. Thomas et al. 1989; Mathis 1990). We sexed salamanders by holding them up to a fibre optic light and using the presence of pigmented testes or eggs as diagnostic for males and females, respectively (Gillette & Peterson 2001). We then placed salamanders in a 15-cm diameter petri dish lined with filter paper and moistened with rainwater. Experiments described below were carried out in these same petri dishes. We changed filter paper once per week, at which time salamanders were fed 15–25 wingless *Drosophila melanogaster*. Feeding was stopped 2 weeks before the start of the experiment to ensure that foragers would be motivated.

Student observers in this experiment consisted primarily of first- and second-year university students at Washington and Lee University (Lexington, Virginia) who were between the ages of 18 and 20 years. Of the 107 students, 49 were male and 58 were female. We refer to male and female students as 'men' and 'women', and hereafter restrict the use of 'male' and 'female' to refer to the sex of the salamanders studied. For the sake of clarity, we use the term 'gender' for men and women, and 'sex' for salamanders.

Students observed salamanders within the petri dishes in laboratory rooms between 1400 and 1700 hours and at temperatures of 20–22°C. All overhead lights were turned off and window blinds were closed to reduce ambient light levels. Red-backed salamanders in nature are primarily active at night and more realistic experiments are carried out in the dark and with less disturbance to focal animals (see below). However, our primary interest was not in the

behaviour of red-backed salamanders, which has been the subject of numerous, detailed behavioural studies (Jaeger & Forester 1993). Rather, we were interested in the consistency of data collection among observers, particularly with respect to the gender and prior expectations of the observer. Thus, we assume only that red-backed salamanders show classifiable behaviours under the conditions of the experiment. It is not necessary for the frequencies of these behaviours to be representative of red-backed salamanders in nature or even in other laboratory experiments.

## Experimental Design

Our study was conducted over 4 days during 18–21 November 2002. Students were divided among five sections that ranged in size from 17 to 24 students. Before arriving at the laboratory, we assigned students a handout that outlined the purpose and procedures of the laboratory. We also provided the students with the following statement about the behaviour of the salamanders, which was intended to give a brief introduction to the topic without biasing their opinions about the possibility of behavioural differences between salamander sexes.

*Red-backed salamanders are a common species throughout the eastern United States and can be found under rocks and logs. We will examine two aspects of salamander behaviour: aggressive behaviour and foraging (food-finding) behaviour. We will then test whether males and females differ in several aspects of foraging behaviour and aggressive behaviour.*

*Why would males and females differ? Females carry eggs, and as such, may have different requirements, preferences and limitations for food and territories as compared to males. Males of course do not have eggs, but are concerned with getting enough food to last them through the winter, when they'll be largely inactive. Red-backed salamanders are known to defend territories (probably they defend individual rocks or logs), though they rarely fight.*

At the beginning of each laboratory period, we described the relevant aggressive behaviours and foraging behaviours to all students as a group. We then randomly divided students into two groups subject to the constraint that men and women were assigned to each group in proportion to their representation within that laboratory section. Each group was then placed in a separate room. We did not tell students why they were separated, only that the salamanders would behave more naturally in a less crowded room.

In each of the five laboratory sections, we designated one of these groups to perform 'blind' trials in which they did not know the sexes of the salamanders until after they had made their observations. The other group performed a 'sighted' trial in which they knew the sexes of their focal animals beforehand. Assignment of the two groups to each of the two laboratory rooms was alternated from day to day. For students in the blind group, salamander sex was marked on the bottom of the petri dish and students recorded this information only after each behavioural trial

was completed. In the sighted trial, salamander sex was labelled on the top of each petri dish and students recorded this information before the trial. Red-backed salamanders show little obvious sexual dimorphism (Quinn & Graves 1999) and it is unlikely that inexperienced observers could distinguish the sexes under the conditions of this experiment. Additionally, students were separated by at least 2 m of desk space and were not permitted to talk during observation periods to preserve the independence of observations. Before initiating observations, students in both groups were asked to make written predictions as to which sex they expected to be more aggressive, to be a more efficient forager, and to be a more active forager. Both groups then completed the following observations in the same manner.

In the first trial, students attempted to determine whether males or females would be more aggressive towards an intruder. To assess aggression, students were asked to record the frequency of five behaviours that may be associated with aggression in plethodontid salamanders. These behaviours were intended to vary in subjectivity, although all have been used in published works on aggressive behaviour in plethodontids (Jaeger 1984; Nishikawa 1987; Lancaster & Jaeger 1995; Rissler et al. 2000). The behaviours were as follows.

- (1) Resident bites intruder ('bites').
- (2) Resident initiates a touch of intruder ('touches').
- (3) Resident moves towards intruder ('moves towards').
- (4) Resident turns to look towards the intruder ('looks towards').
- (5) Resident in all-trunk-raised posture ('ATR'), in which its entire trunk is lifted off the substrate.

For the aggression trial, each student was given two salamanders kept in separate petri dishes. One dish was labelled 'R' for resident, the other was labelled 'I' for intruder. Salamanders were randomly assigned to students subject to the constraint that male and female salamanders were equally divided among men and women. The student picked up the intruder from its dish and placed it in the dish of the resident, on the opposite side from where the resident was located. Students then observed the resident salamander for 20 min divided into 40 intervals of 30 s. Within each interval, students recorded a '+' when any of the aggressive behaviours were observed. If one of the salamanders went under the filter paper in the dish during the 20-min period, the student made a note of the time and stopped making observations until the salamander reappeared or the 20 min were completed.

For the foraging trial, each student was randomly assigned a new salamander again subject to the constraint that male and female salamanders were divided equally among male and female students. As with the aggression trial, one group of students knew the sex, the other did not. At the beginning of the trial, students introduced 10 wingless fruit flies (*D. melanogaster*) into the petri dish containing the focal salamander. In this trial, students were asked to record the following three foraging behaviours.

- (1) Time to eat seven out of 10 introduced flies ('time').
- (2) The number of misses in attempts to capture a fly ('misses'). Salamanders make a quick lunge with their

head when they attempt to capture prey, so we expected that 'misses' could be consistently observed.

(3) The number of steps taken with the front forelimbs during the trial ('steps'). We defined a step as any time a forelimb was lifted and then lowered back to the surface of the petri dish.

For misses and steps, students recorded the total number of behaviours observed until seven of the 10 flies were consumed. If seven out of 10 flies were not consumed within 10 min the trial was terminated at this point. We used the first two measurements as indexes of foraging efficiency with reference to the a priori predictions made by students. We used the number of steps as an index of foraging activity.

We used 153 salamanders in the experiments, and no salamander was used more than once in each trial. Most of the salamanders used in the aggression trials were also used in the foraging trials. For consistency, these salamanders were all used in the aggression trial first, then given at least 48 h before their subsequent use in the foraging trials. Each student ran one aggression trial and one foraging trial.

## Statistical Analysis

We used generalized linear models to evaluate biases in the observation of each of these behaviours, with the exception of bites, which were too infrequent to be analysed statistically. Our original model contained main effects of observer gender, salamander sex and experiment type (blind or sighted). We used the three-way interaction between observer gender, salamander sex and experiment type to evaluate the strength of gender identification bias. Gender identification bias, as defined above, refers to different perceptions of male and female salamanders by men and women that should occur only in the sighted experiments, which defines the three-way interaction term. Two-way interaction terms were considered noninformative with respect to our hypotheses, so were not included in the models. Post hoc inclusion of these terms did not substantially alter any of our results.

We evaluated prediction bias using a separate model because predictions were not experimentally crossed with other main effects. Because some predictions were much more common than others (e.g. 'males will be more aggressive', see Results), not all predictions were adequately represented across combinations of other treatments. Thus, we used a simplified model to evaluate prediction biases that included data only from the sighted experiments. We defined prediction bias as an interaction between the prediction of the observer (e.g. 'females will be more aggressive') and the sex of the salamander on recorded levels of behaviour. This interaction determines whether observers who made a specific prediction about male/female differences actually observed males and females differently than those who made other predictions. This interaction includes effects where observers see what they expect to see, a bias also called 'confirmatory bias' (MacCoun 1998). We also included main effects of observer gender, salamander sex and the prediction made

as covariates but do not present significance tests for these terms since they were previously tested for the full data set as described above (with the exception of the main effect for prediction made, which never approached significance). Other interaction terms were not defined with respect to our hypotheses and were therefore not included in our models. Post hoc inclusion of these terms did not alter any of our conclusions. All analyses were carried out using PROC GENMOD in SAS 8.12. We used type III sums of squares for significance tests in all cases, and we used  $\alpha$  of 0.05 to assess statistical significance. This value for  $\alpha$  does not take into account potential problems with testing multiple behaviours within the same experiment. However, we ultimately deal only with those results that were found to be significant for at least two observed behaviours (see below). The chance of at least two results appearing statistically significant by chance alone is  $1 - P(0 \text{ significant results}) - P(1 \text{ significant result}) = 1 - (0.95)^7 - 7 \times (0.95)^6 \times (0.05)^1 = 0.045$ . Thus, our experiment-wise type I error rate was still less than 0.05, even by these conservative criteria (Moran 2003).

For each model, we used the observer/salamander pair as the experimental unit. This ignores the grouping of students within separate rooms. However, in addition to alternating experiment types among rooms, we also carried out an initial analysis for a main effect of room on each behaviour analysed. The room variable did not approach significance for any of the variables ( $P > 0.15$  in all cases). Thus, we feel it is reasonable to consider each student/salamander combination as a statistically independent unit.

Final data for the four aggression behaviours (touches, moves towards, looks towards and ATR) consisted of the proportion of 30-s periods in which the behaviour was observed. Each of these proportions was arcsine square-root transformed for normality before analysis (Sokal & Rohlf 1995). Two of the foraging behaviours were counts (misses and steps) and were thus treated as Poisson-distributed responses in our linear models. Time (in seconds) for the foraging trials did not differ significantly from a normal distribution, and therefore was not transformed before analysis. Several data points were excluded because of obvious inconsistencies in data entry (e.g. the number of behaviours exceeded the number of time intervals) or because salamanders made no attempts to consume any of the flies.

As part of our analysis of prediction biases, we asked whether men and women differed in the predictions they made about salamander behaviour, independent from the outcome of experimental trials. We used  $G$  tests to ask whether men and women made significantly different predictions about male and female salamanders.

## RESULTS

### Main Analysis

#### Aggression variables

The three-way interaction representing gender identification bias was not significant for any of the four

aggression variables (Table 1, Fig. 1). That is, we found no evidence that men and women viewed male and female salamanders differently in the sighted versus blind treatments. However, several main effects were significant (Table 1). For touches, moves towards and looks towards, the frequencies of recorded behaviours were significantly or marginally significantly higher in the sighted experiments than in the blind experiments. For ATR, men recorded higher overall frequencies of the behaviour than did females across both the sighted and blind experiments (Table 1).

#### Foraging variables

The gender identification interaction was again non-significant for all behaviours analysed (Table 2, Fig. 2). However, as occurred with ATR, the number of misses recorded was significantly higher for men than for women (Table 2). For steps and time, no main effects or interactions were statistically significant (Table 2).

#### Prediction Bias Analysis

The interaction between predictions made and salamander sex did not significantly affect behavioural frequencies for any of the seven variables analysed (Table 3). That is, observers a priori predictions about differences between male and female salamanders did not appear to affect the data collected. For aggression, men were somewhat more likely than women to predict that male salamanders would be more aggressive (Fig. 3a), but this difference was not statistically significant ( $G_1^2 = 1.63$ ,  $P = 0.20$ ). For foraging efficiency, women were significantly more likely to predict that female salamanders would be more efficient foragers ( $G_1^2 = 4.61$ ,  $P = 0.03$ ; Fig. 3b). For foraging activity, men and women were similarly likely to predict that male salamanders would be more active foragers ( $G_1^2 = 0.049$ ,  $P = 0.82$ ; Fig. 3c).

### DISCUSSION

We found no evidence of gender identification biases for any of the aggression or foraging behaviours observed.

**Table 1.** Main analysis of observation biases for aggression variables

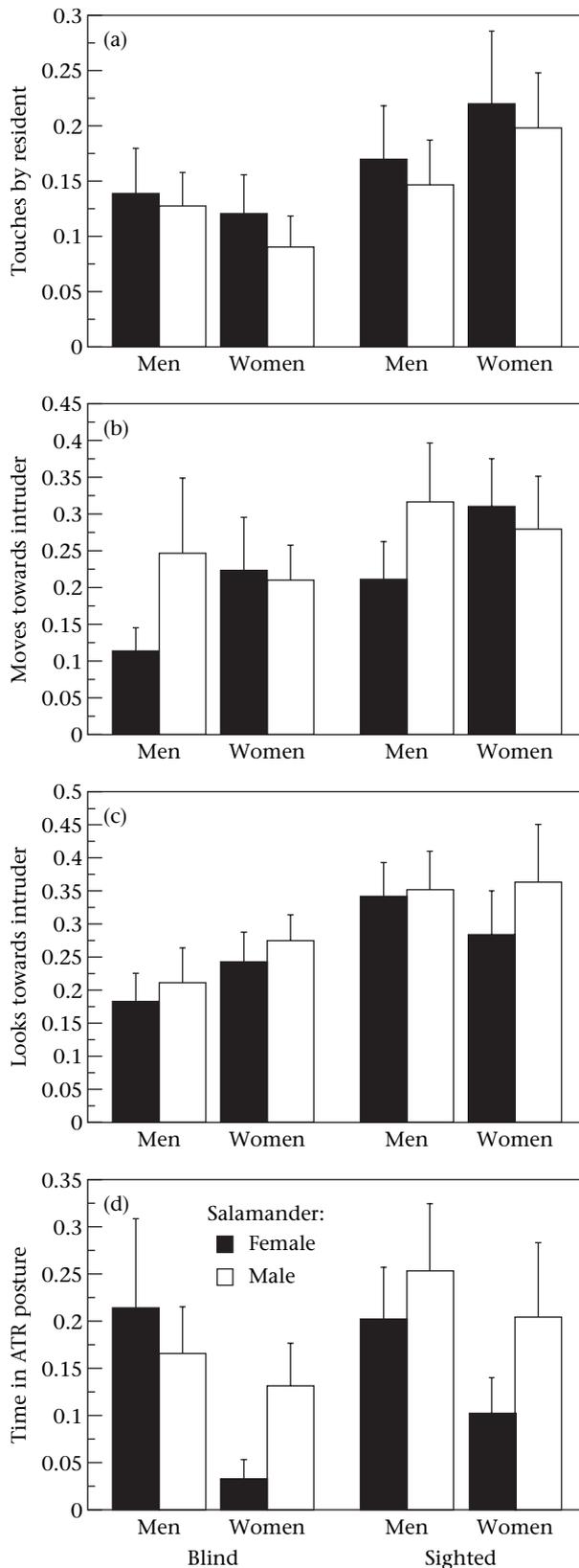
Variable	Observation bias	<i>F</i>	<i>df</i>	<i>P</i>
Touches	Observer gender	0.16	1,97	0.69
	Experiment type	3.95	1,97	0.05**
	Gender identification	0.38	4,97	0.82
Moves towards	Observer gender	0.53	1,97	0.46
	Experiment type	3.02	1,97	0.08*
	Gender identification	0.63	4,97	0.64
Looks towards	Observer gender	0.24	1,97	0.62
	Experiment type	7.15	1,97	0.01**
	Gender identification	0.43	4,97	0.76
ATR	Observer gender	4.49	1,97	0.04**
	Experiment type	1.63	1,97	0.20
	Gender identification	0.47	4,97	0.76

\* $P < 0.10$ ; \*\* $P < 0.05$ .

That is, men and women observers did not appear to observe male and female salamanders differently. We also found no evidence of prediction biases. Although observers expected male and female salamanders to differ with respect to aggression levels and foraging activity and efficiency, there was no evidence that these predictions biased data collection. However, we did find some biases with respect to observers' gender and their knowledge of the sex of the salamanders observed. For time spent in aggressive postures (ATR) and for the number of misses in attempts to capture food, men observed greater frequencies of these behaviours than did women, regardless of the sex of the salamander observed. Additionally, for three measures of aggression (moves towards intruder, touches intruder, looks towards intruder), observers recorded higher levels of behaviour in the sighted experiments, in which they knew the sex of the salamander, than they did in the blind experiments, in which the sex of the salamander was not known until after data collection. These biases were found for both men and women observers. For the last two behavioural indexes, total time spent foraging and number of steps taken while foraging, no statistically significant observation biases were found.

These results are surprising in that biases that have been suggested in the philosophy of science literature (i.e. gender identification and prediction) were not found at all, whereas other, more subtle biases were found for at least two of the seven behaviours observed. From these results, it is evident that blind experiments, in which observers are unaware of the treatment groups, may not be sufficient to eliminate all observation biases. Gender identification and prediction biases, to the extent that they exist, should be reliably eliminated when observers are not aware of which samples belong to which treatment groups. However, overall differences in data collected between men and women observers may not be reduced with blind experiments. Similarly, differences in observed frequencies of behaviours between blind and sighted experiments will not necessarily be remedied by conducting only blind experiments.

The behaviours observed varied in their susceptibility to subjective interpretation. We initially predicted that time spent in ATR would be the most subjective behaviour for inexperienced observers, whereas we predicted that time to complete the foraging trial would be the least subjective. These predictions were generally supported by the overall effects of observer gender on observed ATR frequency and by the lack of any significant biases for foraging time. For the range of other behaviours, we found it difficult to predict in advance which would and would not be biased. Even in retrospect, it is not obvious why 'steps' showed no biases, whereas 'touches' showed an overall bias between the sighted and blind experiments. Observer bias could differ between behaviours scored by one-zero sampling (e.g. touches, moves towards, looks towards and ATR) and behaviours scored by continuous sampling (e.g. misses and steps). Three of the one-zero behaviours differed in recorded frequency between blind and unblind trials, whereas none of the continuously scored behaviours differed in this manner. However, these differences might also reflect different propensities for bias between measures of aggression and



**Figure 1.** Observation biases for aggression variables ( $\bar{x} \pm SE$ ). Proportion of observation intervals including (a) touches of intruders initiated by residents, (b) moves by residents towards intruders, (c) looks by residents towards intruders and (d) the all-trunk-raised posture (ATR) by residents.

**Table 2.** Main analysis of observation biases for foraging variables

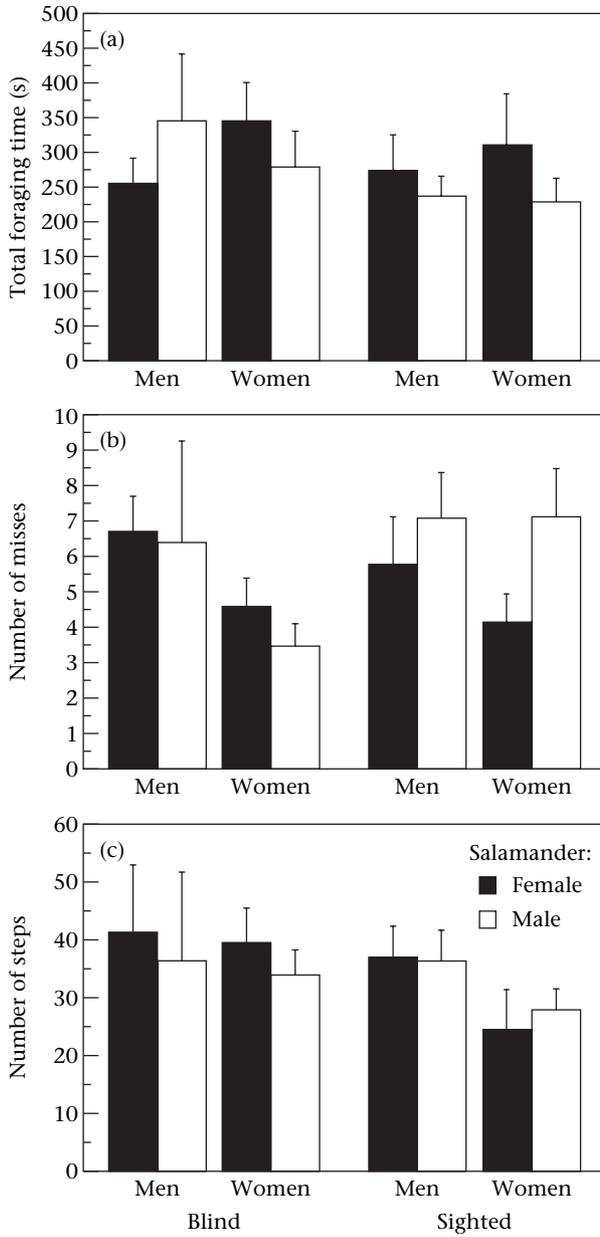
Variable	Observation bias	F	df	P
Time	Observer gender	0.01	1,85	0.97
	Experiment type	1.94	1,85	0.19
	Gender identification	0.11	4,85	0.89
Misses	Observer gender	3.97	1,85	0.05*
	Experiment type	0.43	1,85	0.65
	Gender identification	0.90	4,85	0.47
Steps	Observer gender	1.32	1,85	0.25
	Experiment type	1.27	1,85	0.26
	Gender identification	0.60	1,85	0.66

\**P* < 0.05.

measures of foraging. The difficulty of predicting sources of observation bias suggests that while intuition might serve as a partial guide to identifying reliable behaviours, more formal studies may be desirable to confirm that behaviours can be consistently recorded across observers.

The most obvious issue with this research is the extent to which the observation biases of undergraduate science students are similar to the biases of practicing researchers. University students are often involved in collecting data in animal behaviour and other fields of organismal biology. However, these students are generally trained as observers and therefore have more experience than the students used in our experiments. Graduate students in animal behaviour presumably acquire even greater observation skills. Thus, it is possible that the biases we observed would be less prevalent in practising scientists than in our sample of university students. At a minimum then, our work shows the importance of carefully training scientific observers rather than simply assuming that observers can collect data in an unbiased way. Alternatively, it is possible that experienced scientists might be more susceptible than inexperienced observers to some kinds of biases. For example, experienced scientists are under greater pressure to publish and might have greater investment in particular outcomes of their research. This could make scientists more subject to prediction or confirmatory bias than the students in our study. Unfortunately, it would be difficult to replicate a study such as this across a large number of practising scientists. A more promising approach would be meta-analyses of published studies that ask similar questions, but use different types of observation methods (e.g. blind versus sighted experiments). In addition, it would be possible to study the effects of observer training on observation bias directly in an experiment similar to ours.

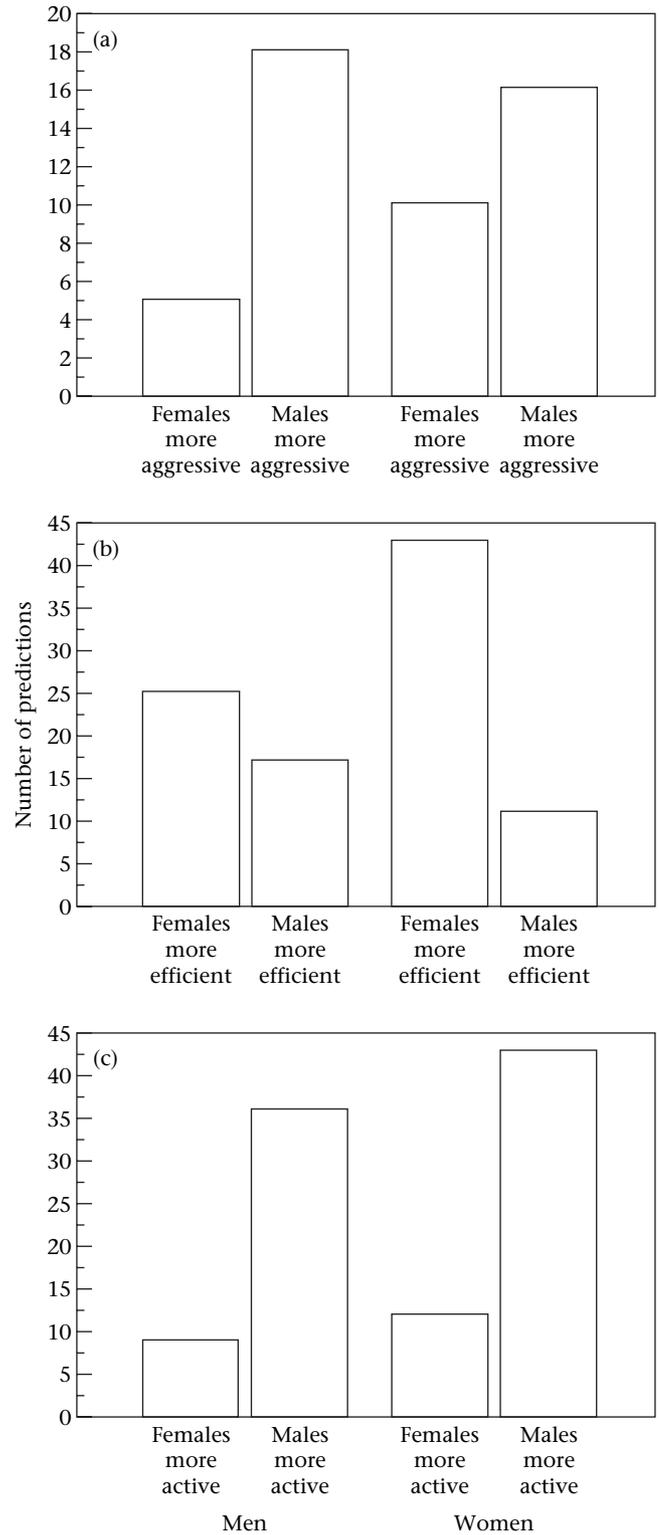
A further issue is that our experiments only identify bias relative to absolute levels of behaviour. Most hypothesis testing in animal behaviour concerns differences in behaviour between experimental groups. Differences between men and women in the frequency of recorded behaviours do not necessarily imply that men and women would reach different conclusions in testing hypotheses. The same applies for the observed biases with respect to blind versus sighted experiments. In order to affect the outcome of hypothesis testing, the magnitude of observer biases would have to vary between treatment groups. This



**Figure 2.** Observation biases for foraging variables ( $\bar{X} \pm SE$ ). (a) Total time to capture seven of 10 flies. (b) Number of misses in attempts to capture flies over the foraging interval. (c) Number of steps taken with the forelimbs during the foraging interval.

**Table 3.** Prediction biases for aggression and foraging variables

Variable	F	df	P
Touches	0.26	1,44	0.61
Moves towards	0.02	1,44	0.90
Looks towards	2.36	1,44	0.13
ATR	0.21	1,44	0.65
Foraging time	1.16	1,37	0.29
Misses	0.62	1,37	0.44
Steps	0.78	1,37	0.38



**Figure 3.** Predictions of observers for male/female differences in salamander behaviour. Predictions for (a) aggressive behaviour, (b) foraging efficiency (total foraging time and number of misses) and (c) foraging activity (number of steps taken during the foraging interval).

scenario may be less likely than that in which overall levels of behaviour vary with gender or experiment type. Indeed, the lack of significant gender identification biases shows that men and women would have reached similar conclusions about the behavioural differences between male and female salamanders within the conditions of our study. However, given some of the more subtle biases found here, the potential for interactions between observer biases and experimental treatments certainly warrants further attention.

A final issue with our work is that gender identification biases in particular, and anthropomorphism in general, may be most likely with animals that are closely related to humans (Zuk 2002). For example, there have been numerous anecdotal accounts of gender biases in the study of primatology (Hrdy 1981; Zuk 2002). Because the behaviour of red-backed salamanders does not closely resemble human behaviour, observations of these animals may be less likely to show gender identification biases than would observations of other species. Nevertheless, we did find some differences between how men and women observed the behaviour of red-backed salamanders. These kinds of biases could be even more of an issue with primates or other mammals.

The main importance of our study is the demonstration that observer biases are potentially an issue in many studies of animal behaviour, and that the nature of these biases can be elucidated with careful studies of observers. We also note that the different recorded behaviours were quite variable in terms of observation biases. Several behavioural observations were affected by observer gender, several were affected by experiment type, and several showed no significant biases at all. This implies that preliminary screening for intra- and interobserver repeatability could be a useful tool in selecting behaviours for more detailed study. Finally, we believe that the potential for observation biases supports the use of blind experiments whenever possible. As our results show, use of blind experiments does not by itself eliminate all possible types of observer biases. However, blind experiments in concert with preliminary tests for repeatability within and between observers could eliminate most kinds of bias. Although these sorts of steps may be time consuming, they could substantially improve the consistency and scientific value of behavioural research, particularly for long-term research programmes.

### Acknowledgments

We thank the autumn 2002 Animal Behavior class and the General Biology laboratory assistants for their help in coordinating and carrying out this experiment. Earlier drafts of the manuscript were improved by comments from Robert Page and two anonymous referees. This work was approved by the Institutional Review Board at Washington and Lee University. Animal care and research methods were approved under IACUC protocol 0402a. Salamanders were collected with the permission of the Virginia Department of Game and Inland Fisheries, permit 021066.

### References

- Altmann, J. 1974. Observational study of behaviour: sampling methods. *Behaviour*, **49**, 227–267.
- Barlow, G. W. 1968. Ethological units of behaviour. In: *The Central Nervous System and Fish Behaviour* (Ed. by D. Ingle), pp. 217–232. Chicago: University of Chicago Press.
- Behar, P. M. & Todd, N. W. 1999. Gender and intra-observer agreement about laryngoscopy of papilloma. *International Journal of Pediatric Otorhinolaryngology*, **50**, 125–131.
- Caine, N. G. 1990. Unrecognized anti-predator behaviour can bias observational data. *Animal Behaviour*, **39**, 195–197.
- David, R. S. & Jaeger, R. G. 1981. Prey location through chemical cues by a terrestrial salamander. *Copeia*, **1981**, 435–440.
- Davidson, K., MacGregor, M. W., MacLean, D. R., McDermott, N., Farquharson, J. & Chaplin, W. F. 1996. Coder gender and potential for hostility ratings. *Health Psychology*, **15**, 298–302.
- Gillette, J. R. & Peterson, M. G. 2001. The benefits of transparency: candling as a simple method for determining sex in red-backed salamanders (*Plethodon cinereus*). *Herpetological Review*, **32**, 233–235.
- Gowaty, P. A. 1997. *Feminism and Evolutionary Biology*. New York: Chapman & Hall.
- Holmes, D. J. & Hitchcock, C. L. 1997. A feeling for the organism? An empirical look at gender and research choices of animal behaviorists. In: *Feminism and Evolutionary Biology* (Ed. by P. A. Gowaty), pp. 184–204. New York: Chapman & Hall.
- Hrdy, S. B. 1981. *The Woman That Never Evolved*. Cambridge: Harvard University Press.
- Jaeger, R. G. 1984. Agonistic behavior of the red-backed salamander. *Copeia*, **1984**, 309–314.
- Jaeger, R. G., Joseph, R. G. & Barnard, D. E. 1981. Foraging tactics of a terrestrial salamander: sustained yield in territories. *Animal Behaviour*, **29**, 1100–1105.
- Jaeger, R. G., Barnard, D. E. & Joseph, R. G. 1982. Foraging tactics of a terrestrial salamander: assessing prey density. *American Naturalist*, **119**, 885–890.
- Jaeger, R. G. & Forester, D. C. 1993. Social behavior of plethodontid salamanders. *Herpetologica*, **49**, 163–175.
- Jordan, R. H. & Burghardt, G. M. 1986. Employing an ethogram to detect reactivity of black bears to the presence of humans. *Ethology*, **73**, 89–115.
- Lancaster, D. L. & Jaeger, R. G. 1995. Rules of engagement for adult salamanders in territorial conflicts with heterospecific juveniles. *Behavioral Ecology and Sociobiology*, **37**, 25–29.
- van Lawick-Goodall, J. 1968. The behaviour of free-living chimpanzees in the Gombe Stream Reserve. *Animal Behaviour Monographs*, **1**, 161–311.
- Lawton, M. F., Gartska, W. R. & Hanks, J. C. 1997. The mask of theory and the face of nature. In: *Feminism and Evolutionary Biology* (Ed. by P. A. Gowaty), pp. 63–85. New York: Chapman & Hall.
- MacCoun, R. J. 1998. Biases in the interpretation and use of research results. *Annual Review of Psychology*, **49**, 259–287.
- Mathis, A. 1990. Territoriality in a terrestrial salamander: the influence of resource quality and body size. *Behaviour*, **112**, 162–174.
- Moran, M. D. 2003. Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, **100**, 403–405.
- Morgan, C. L. 1890. *Animal Life and Intelligence*. London: E. Arnold.
- Nishikawa, K. C. 1987. Interspecific aggressive behaviour in salamanders: specific-specific interference or misidentification. *Animal Behaviour*, **35**, 263–270.
- Ollason, J. C. & Dunnet, G. M. 1980. Nest failures in the fulmar: the effect of observers. *Journal of Field Ornithology*, **51**, 39–54.

- Pierotti, R., Annett, C. A. & Hand, J. L. 1997. Male and female perceptions of pair-bond dynamics: monogamy in western gulls, *Larus occidentalis*. In: *Feminism and Evolutionary Biology* (Ed. by P. A. Gowaty), pp. 261–275. New York: Chapman & Hall.
- Quinn, V. S. & Graves, B. M. 1999. A technique for sexing red-backed salamanders (*Plethodon cinereus*). *Herpetological Review*, **30**, 32.
- Rasmussen, D. R. 1979. Correlates of patterns of range use of a troop of yellow baboons (*Papio cynocephalus*). I. Sleeping sites, impregnable females, births, and male emigrations and immigrations. *Animal Behaviour*, **27**, 1098–1112.
- Rissler, L. J., Barber, A. M. & Wilbur, H. M. 2000. Spatial and behavioural interactions between a native and introduced salamander species. *Behavioral Ecology and Sociobiology*, **48**, 61–68.
- Robinson, M. E. & Wise, E. A. 2003. Gender bias in the observation of experimental pain. *Pain*, **104**, 259–264.
- Rosenthal, R. 1966. *Experimenter Effects in Behavioural Research*. New York: Appleton-Century-Crofts.
- Sokal, R. R. & Rohlf, F. J. 1995. *Biometry*. New York: W.H. Freeman.
- Thomas, J. S., Jaeger, R. G. & Horne, E. A. 1989. Are all females welcome? Agonistic behavior of male red-backed salamanders. *Copeia*, **1989**, 915–920.
- Thorndike, E. L. 1898. Animals' intelligence: an experimental study of the associative processes in animals. *Psychological Review*, **20**, 158–177.
- Washburn, M. F. 1908. *The Animal Mind*. New York: Macmillan.
- Watson, J. B. 1924. *Behaviourism*. New York: Norton.
- Zuk, M. 2002. *Sexual Selections*. Berkeley: University of California Press.