Experimenter expectancy bias does not explain Eurasian jays' (*Garrulus glandarius*) performance in a desire-state attribution task

Ljerka Ostojić[1], Edward W. Legg[1], Arne Dits[2], Natalie Williams[1], Katharina F. Brecht[1], Michael Mendl[3] and Nicola S. Clayton[1]

[1]University of Cambridge, Department of Psychology, Downing Street, Cambridge, CB2 3EB, UK

[2]University of Amsterdam, Institute for Biodiversity and Ecosystem Dynamics, Amsterdam, Science Park 904, NL

[3]Centre for Behavioural Biology, School of Veterinary Science, University of Bristol, Langford House, Langford, Bristol, BS40 5DU, UK

Corresponding author: Ljerka Ostojić, Email: lo245@cam.ac.uk

Abstract

Male Eurasian jays have been found to adjust the type of food they share with their female partner after seeing her eat one type of food to satiety. One interpretation of this behavior is that the male encoded the female's decreased desire for the food she was sated on, and adjusted his behavior accordingly. However, in these studies, the male's actions were scored by experimenters who knew on which food the female was sated. Thus, it is possible that the experimenters' expectations (sub-consciously) affected their behavior during tests that, in turn, inadvertently could have influenced the males' actions. Here, we repeated the original test with an experimenter who was blind to the food on which the female was sated. This procedure yielded the same results as the original studies: the male shared food with the female that was in line with her current desire. Thus, our results rule out the possibility that the Eurasian jay males' actions in the food sharing task could be explained by the effects of an experimenter expectancy bias.

Keywords: corvids; mental state attribution; food sharing; experimenter expectancy bias

Introduction

A common criticism of research in animal behavior is that many studies do not attempt to prevent the influence of the experimenter's expectations on the reported results (Beran, 2012; Burghardt et al., 2012; Kardish et al. 2015; Sebeok & Umiker-Sebeok, 1980). The issue is that whenever an animal's behavior is scored by an experimenter who is not blind to the testing conditions, the results are susceptible to the experimenter's expectations. For example, an animal's action might be directly affected by the experimenter's conscious or unconscious behaviors, or the experimenter might interpret the animal's action to match how they expect the animal to behave in the test situation. These experimenter expectancy biases have been acknowledged for over 100 years (Pfungst, 1911; Rosenthal, 1976), yet very few contemporary studies in the field of animal behavior involve blind experimenters (Burghardt et al., 2012).

Two recent studies suggested that Eurasian jays (*Garrulus glandarius*) might be capable of desire-state attribution. Male Eurasian jays were shown to be sensitive to their female partner's current desire when sharing food with her during the breeding season (Ostojić et al., 2013; 2014). After seeing her eat one particular food to satiety, the male subsequently adjusted his sharing behavior in a way that matched the female's decreased desire for the food on which she was sated. In these studies, an experimenter was present at the time of testing. This experimenter first gave a particular food to the female during the pre-feeding phase and then offered the male the test foods, by holding a different type of food in each hand and live scored which food the male chose and which food he then shared with his female partner. Thus, the experimenter was knowledgeable about the different pre-feeding treatments in the experiment when they were carrying out the food sharing test.

In this set-up, an experimenter expectancy bias could theoretically influence the relevant measurement – how much of each type of food the male shares in the different pre-feeding conditions—in three different ways. Firstly, the experimenter's behavior could bias which food the male takes from the experimenter, which in turn might influence what food the male shares with the female. This type of bias is unlikely because the food chosen by the males does not differ depending on what food the female was pre-fed (Ostojić et al., 2014). Notably, although the male chooses a similar pattern of food across the different pre-feeding trials, what he shares differs between the trials. This is because, apart from sharing the food with the female, the male can also eat the chosen food himself or cache it. Secondly, the experimenter's expectation could influence their scoring of the male's actions. This type of bias is unlikely to affect the results because inter-observer reliability between an experimenter and a naive rater, obtained when the food shared was scored from videos, was consistently high (Cohen's κ =.87 in Ostojić et al., 2013, and Cohen's κ = .82 in Ostojić et al., 2014). Finally, the experimenter's behavior might affect when and what the male shares with the female. When an experimenter needs to be present during the test phase, the only way to address this issue is for this experimenter to be blind to the testing conditions. In this case the experimenter who offers the food to the male and scores the male's behavior would need to be ignorant of what food the female has been pre-fed. Importantly, if the original results could be reproduced using a blind experimenter, this would provide evidence against all three ways in which an experimenter's expectation could have influenced the original data.

In the current study, we repeated the main test from the original study, in which the male saw the female being pre-fed and subsequently could share the test foods with her ('seen' condition; Ostojić et al., 2013). However, this time the birds were tested by two experimenters. One experimenter conducted the pre-feeding phase and thus knew what food the female would have desired on the different testing days (henceforth the knowledgeable experimenter). Another experimenter, who had no knowledge of what food the female had been pre-fed (henceforth the blind experimenter), presented food to the male and scored his behavior during the food sharing test phase. If the previous findings that the male shared food in accordance with the female's specific satiety were merely an artefact of an experimenter expectancy bias, then the sharing pattern scored by the blind experimenter should either not change between the different pre-feeding conditions or show a pattern that is not in accordance with the female's specific satiety. In contrast, if the previous findings rely on the male's ability to cater to the female's desire, then the sharing pattern scored by the blind experimenter should exhibit the original effect and be in line with the female's specific satiety.

Methods

Subjects

Eight male and female Eurasian jay pairs were tested during the breeding season (March to June) in 2015, which is the only time when jays share food. All birds first participated in a specific satiety experiment (for details of procedure see Ostojić et al., 2013), which ensured that they had specific satiety to the test foods. Pairs included 16 jays from two colonies (colony 1: n = 8, all 8 years old; colony 2: n = 8, all 7 years old). The two colonies were housed in two separate outdoor aviaries (20 x 6 x 3 m) and tested in indoor testing compartments (2 x 1 x 2 m). The birds could access the indoor compartments from the aviary via opaque trap doors (0.5 x 0.5 m), which were opened and closed by the experimenter. Birds had ad libitum access to water and outside of testing were fed a maintenance diet of soaked dog biscuits, cheese, seeds, nuts and fruit. The study was approved by the University of Cambridge Ethics Review Process.

Procedure

To ensure that the birds were mildly hungry and thus motivated to eat the pre-feeding food, the birds' maintenance diet was removed approximately 2 h before testing. All pairs were tested only once a day. During testing, males and females were called into separate, adjacent indoor compartments that were joined by a wire mesh window.

All trials consisted of a pre-feeding and a test phase. For colony 1, KFB served as the experimenter who conducted the pre-feeding phase (knowledgeable experimenter) and LO served as the experimenter who conducted the test phase (blind experimenter). For colony 2, NW served as the knowledgeable experimenter and EWL served as the blind experimenter. During the pre-feeding phase, the knowledgeable experimenter pre-fed the female different foods (a handful of maintenance diet – MD, 50 wax moth larvae – W, or 50 mealworm beetle larvae – M) and the male with MD on all three trials. During this phase the jays were prevented from sharing food with each other by a transparent Perspex barrier that was attached to the mesh between the male's and the female's compartments. At the end of the 15-minute long pre-feeding phase, the knowledgeable experimenter removed all foods from the testing compartments and removed the Perspex barrier. The pre-feeding food was prepared and counted by the knowledgeable experimenter out of sight of the blind experimenter. During the subsequent test phase, the blind experimenter gave the males 20 choices between a single W and a single M. For six males, the experimenter held one larva in each hand against the mesh of the compartment. For three males who were not tame enough for this procedure (Ayton, Dublin, Lisbon), the choices were presented on a platform inside the compartment. The position of the food was pseudo-randomised with no food appearing on the same side for more than two consecutive trials. If no choice was made within 30 s, the foods were removed. Each opportunity to make a choice was followed by 40 s, in which males could either eat, cache or feed the food to the female through the mesh.

All pairs started with an 'informed' baseline, namely a trial in which the female was pre-fed maintenance diet (MD) and which was known to both experimenters. The aim of this 'informed' baseline was to test whether the birds were comfortable enough with the procedure of two experimenters testing them. To proceed to testing, the males had to choose at least 10 of the 20 choices and share food with their female partner at least twice. Each pair was given a maximum of five 'informed' baselines. Subsequently, birds received three trials (female pre-fed MD – baseline, female pre-fed W or female pre-fed M), the order of which was randomised for each pair by the knowledgeable experimenter and was unknown to the blind experimenter.

Analysis

Data were live scored by LO for colony 1 and EWL for colony 2. The results from the baseline (female pre-fed MD) showed that males preferred to choose and share W over M (Table 1a). Following the analysis of Ostojić et al. (2014), to investigate how the female's specific satiety to the two test foods affected this preference, for each trial, we calculated the number of W minus the number of M chosen or shared: (W-M). This difference score accounts for males whose preference for W over M is so high that they only ever share W with the female. For these males, a response to the female's specific satiety is possible by sharing a different number of W in the test trials (see Ostojić et al., 2014).

Graphs show the difference between these values in a test trial (female pre-fed W or M) and the baseline (female pre-fed MD): $[(\text{W-M})_{\text{female pre-fed W or M}} - (\text{W-M})_{\text{female pre-fed MD}}]$. This ensured that individual variation in the amount of food shared as well as in general food preferences were taken into account. If the male could take the female's specific satiety into account, in a direct comparison between the two test trials (female pre-fed W and female pre-fed M) his preference for W over M relative to the baseline was expected to be lower after the female had been pre-fed W than after the female had been pre-fed M.

To test whether an experimenter expectancy bias might influence the magnitude of the effect, we further compared the data from the current study to the data obtained in the original food-sharing test (Ostojić et al., 2013; 2014), in which the trials were score by experimenters who were knowledgeable about what food the female had been eating during the pre-feeding phase. For these analyses we compared the pattern of items chosen/shared (i.e., the difference of the difference score between the two test trials) in the current study with the pattern of items chosen/shared (i.e., the difference of the difference score between the two test trials) in the original study.

In the original study, the measurement used to investigate the males' sharing pattern was the proportion of W out of total number of worms shared (see Ostojić et al., 2013). In contrast to this original study, in the current study some males shared only W across all test trials, such that a response to the female's specific satiety was only possible by modifying the number of W shared with her. Thus, instead of proportions, we used the difference score of number of W minus number of M as explained above. Consequently, it was necessary to re-analyse the original data, not just for the 'seen' condition, which was directly compared to the data obtained in the current study, but also for the 'unseen' condition as reported in Ostojić et al. (2013).

In both cases, when we conducted the analyses using the difference scores instead of proportions we found the same results as reported in the original study. In the 'seen' condition, the males catered for the female's specific satiety by showing a smaller preference for sharing W over M relative to the baseline when the female was pre-fed W than when she was pre-fed M (Z=2.45, p=.007). In the 'unseen' condition, the males did not alter their sharing behavior across the test trials (Z=-0.85, p=.312) and this sharing pattern differed from that exhibited in the 'seen' condition (Z=-2.01, p=.031).

All analyses were planned contrasts, performed using exact permutation tests (Anderson, 2001). All tests were one-tailed. Alpha was set at.05.

Results

All pairs except one passed the 'informed' baseline on their first trial. This pair did not pass the required criteria within five trials and thus could not subsequently be tested (male: Ayton). Although they passed the 'informed' baseline, another pair did not share anything in the three test trials, which was possibly due to the weather conditions when this pair was tested (male: Pendleton). The testing compartments were very hot and this might have decreased the birds' motivation to engage in food sharing. Thus, only the data from the remaining six pairs could be included in the analyses (and are shown in Table 1).

The female's specific satiety affected the male's sharing pattern: the male's preference for sharing W over M relative to the baseline was lower after the female had been pre-fed W than after she had been pre-fed M (n = 6, Z = -1.69, p =.031, Cohen's d = 0.87; Figure 1a, raw data see Table 1a). In contrast, the female's specific satiety did not affect the male's choices of the two foods: the male's preference for choosing W over M relative to the baseline did not differ whether the female had been pre-fed W or M ($n = 6$, $Z = -1.34$, $p = .187$, Cohen's $d = 0.60$; Figure 1b, raw data see Table 1b).

In addition, the males' behaviors as scored by the blind experimenters did not differ from the data reported in the original studies (Ostojić et al., 2013; 2014; raw data presented in Table 2), in which the experimenters knew which food the female had been pre-fed (items shared: $n = 6$, $Z = -0.61$, $p = .750$, Cohen's $d = 0.23$; items chosen: $n = 6$, $Z = 1.06$, $p = .844$, Cohen's $d = 0.44$).

Discussion

The male Eurasian jays adjusted the food shared with their female partner according to what food they saw her eat before the sharing event. Specifically, the male jays responded to the change in the female's specific satiety and thus decreased desire for the pre-fed food. Critically, in the current study, the food shared by the male was live scored by experimenters who were blind to the testing condition, i.e., to what food the female had been pre-fed and on which she had thus been sated. In addition, the male's sharing pattern did not differ from the one shown in previous studies, in which the male's behavior was scored by knowledgeable experimenters (Ostojić et al., 2013; 2014), suggesting that the magnitude of the effect did not differ between the studies. Thus, the current findings provide evidence that an experimenter expectancy bias cannot explain the male's sharing pattern.

In contrast to the male's sharing pattern, the food chosen by the male did not respond to the female's specific satiety. The same result was found in previous studies, in which the experimenter was not blind to the testing conditions. Thus, although previous results indicated that an experimenter expectancy bias was unlikely to explain the male's decision as to what food to take, the current findings provide further evidence that the male's choices are not influenced by the experimenter's expectations.

To ensure that the relevant experimenter is blind to the testing conditions required that two experimenters tested a particular population of jays. Although it has been claimed that introducing blind experimenters would be straightforward in behavioural tests (Kardish et al., 2015), this procedure is not trivial and often constrained by serious practical concerns. Corvids are neophobic and their performance in a cognitive task is affected by the level of familiarity with the experimenter (Cibulski et al., 2013). Consequently, it is crucial that the birds are familiar with both experimenters, which requires a large time investment on the part of an experimenter who does not usually work with that particular colony of birds. In addition, the involvement of two experimenters might increase the demands on the birds' attention and thus interfere with other experimental manipulations. If birds are required to attend to critical experimental manipulations, then a change in experimenters might result in either proactive or retroactive interference, potentially skewing the obtained data (Grant, 1988; Maki et al., 1977). By overcoming these issues in the current study we provide evidence against an experimenter expectancy bias in the food-sharing task, thus ensuring that that the males' actions can be interpreted as a consequence of the manipulations of the female's desire.

References

Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences*, *58*, 626-639. doi:10.1139/f01-004.

Balph, D. F., & Balph, M. H. (1983). On the psychology of watching birds: the problem of observer-expectancy bias. *Auk, 100*, 755-757.

Beran, M. J. (2012). Did you ever hear the one about the horse that could count? *Frontiers in Psychology, 3,* 1-2. doi:10.3389/fpsyg.2012.00357

Burghardt, G. M., Bartmess-LeVasseur, J. N.,Browning, S. A., Morrison, K. E., Stec, C. L., Zachau, C. E., & Freeberg, T. M. (2012). Perspectives – Minimising observer bias in behavioural studies: a review and recommendations. *Ethology*, *118*, 511-517.

Cibulski, L., Wascher, C. A. F., Weiß, B. W., & Kotrschal, K. (2013). Familiarity with the experimenter influences the performance of Common ravens (*Corvus corax*) and Carrion crows (*Corvus corone corone*) in cognitive tasks. *Behavioural Processes*, *103*, 129-137. doi:10.1016/j.beproc.2013.11.013

Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42-45.

Grant, D. S. (1988). Sources of visual interference in delayed matching-to-sample with pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 368-375.

Kardish, M. R., Mueller, U. G., Amador-Vargas, S., Dietrich, E. I., Ma, R., Barrett, B., & Fang, C. (2015). Blind trust in unblended observation in ecology, evolution and behavior. *Frontiers in Ecology and Evolution*, *3*, 1-4. doi:10.3389/fevo.2015.00051

Maki, W. S., Moe, J. C., & Bierley, C. M. (1977). Short-term memory for stimuli, responses, and reinforcers. *Journal of Experimental Psychology: Animal Behavior Processes*, *3*, 156-177.

Ostojić, L., Shaw, R. C., Cheke, L. G., & Clayton, N. S. (2013). Evidence suggesting that desire-state attribution may govern food sharing in Eurasian jays. *Proceedings of the National Academy of Science USA*, *110*, 4123-4128. doi:10.1073/pnas.1209926110.

Ostojić, L., Legg, E. W., Shaw, R. C., Cheke, L. G., Mendl, M., & Clayton, N. S. (2014). Can male Eurasian jays disengage from their own current desire to feed the female what she wants? *Biology Letters*, *10,* 20144042. doi:10.1098/rsbl.2014.0042.

Pfungst, O. (1911). *Clever Hans. The horse of Mr. von Osten.* New York, NY: Henry Holt.

Rosenthal, R. (1976). *Experimenter effects in behavioural research*. New York, NY: Halstad Press.

Sebeok, T. A. & Umiker-Sebeok J. (1980). *Speaking of apes: A critical anthology of two-way communication of man*. New York, NY: Plenum Press.

Acknowledgements

We thank the BBSRC, the ERC and the Leverhulme Trust for funding the research. AD was funded by an ERASMUS scholarship and KFB was funded by Lucy Cavendish College and the Cambridge Commonwealth and European Trust.

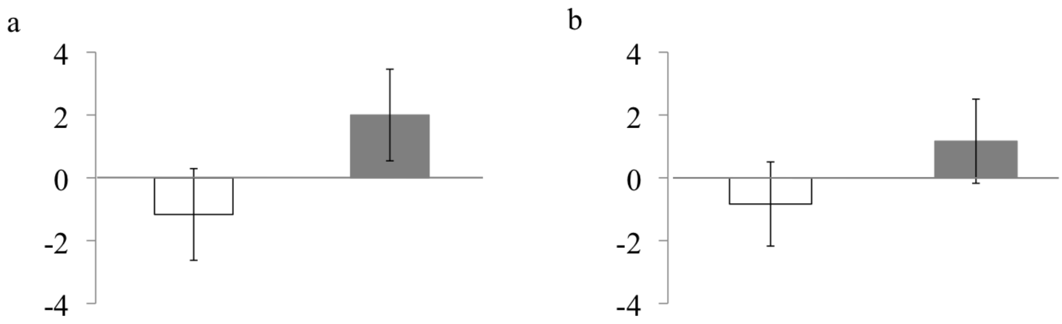

Figure 1. Mean difference in the number of W minus the number of M (a) shared and (b) chosen between the pre-fed W and the pre-fed MD trials (white bars) and between the pre-fed M and the pre-fed MD trials (grey bars). Values under zero denote a decrease in the preference for W over M relative to the baseline (pre-fed MD) and values over zero denote an increase in the preference for W over M relative to the baseline (pre-fed MD). Error bars denote 95% confidence intervals.

Table 1: Items shared and chosen by each male as scored by the blind experimenters

	Items shared						Items chosen					
Pre-fed:	MD		W		M		MD		W		M	
	W	M	W	M	W	M	W	M	W	M	W	M
Caracas	3	1	4	0	5	0	17	3	19	1	20	0
Lima	8	0	3	0	13	0	17	3	16	4	19	1
Dublin	9	3	8	1	11	0	17	3	19	1	18	2
Lisbon	7	1	6	1	9	1	16	4	13	7	13	7
Romero	5	2	1	0	5	4	11	9	8	12	11	9
Hoy	5	2	1	0	2	0	13	6	14	6	14	6

MD = maintenance diet, W = wax moth larvae, M = mealworm beetle larvae

The row 'pre-fed' refers to the food that was given to the female during the pre-feeding phase.

The data from two additional males (Ayton, Pendleton) are not shown in the table and were not included in the analysis. Ayton did not share anything in the pre-test ('informed' baseline) and thus did not participate in the main test. Pendleton passed the pre-test but did not share any food with his female partner in any of the three trials of the main test.

Table 2: Items shared and chosen by each male in the 'seen' condition of the original study, in which the male's behavior was scored by knowledgeable experimenters

	Items shared						Items chosen					
Pre-fed:	MD		W		M		MD		W		M	
	W	M	W	M	W	M	W	M	W	M	W	M
Caracas	5	2	1	2	7	1	15	5	8	12	19	1
Lima	6	3	1	0	4	1	12	8	15	5	19	1
Dublin	11	2	6	1	11	0	18	2	19	1	17	3
Lisbon	7	0	3	2	6	0	13	2	8	9	15	1
Romero*	6	0	2	1	7	1	14	6	10	6	11	9
Hoy*	5	0	4	0	5	0	11	6	10	4	12	5

MD = maintenance diet, W = wax moth larvae, M = mealworm beetle larvae

The row 'pre-fed' refers to the food that was given to the female during the pre-feeding phase.

* denotes males that did not take part in the Ostojić et al. (2013) study but which have been tested on the 'seen' condition of the original test as part of the Ostojić et al. (2014) study.