

Journal of Comparative Psychology

Replication of the Mirror Mark Test Experiment in the Magpie (*Pica pica*) Does Not Provide Evidence of Self-Recognition

Manuel Soler, José Manuel Colmenero, Tomás Pérez-Contreras, and Juan Manuel Peralta-Sánchez
Online First Publication, May 14, 2020. <http://dx.doi.org/10.1037/com0000223>

CITATION

Soler, M., Colmenero, J. M., Pérez-Contreras, T., & Peralta-Sánchez, J. M. (2020, May 14). Replication of the Mirror Mark Test Experiment in the Magpie (*Pica pica*) Does Not Provide Evidence of Self-Recognition. *Journal of Comparative Psychology*. Advance online publication. <http://dx.doi.org/10.1037/com0000223>

Replication of the Mirror Mark Test Experiment in the Magpie (*Pica pica*) Does Not Provide Evidence of Self-Recognition

Manuel Soler, José Manuel Colmenero,
and Tomás Pérez-Contreras
University of Granada

Juan Manuel Peralta-Sánchez
University of Granada and Doñana Biological Station,
Seville, Spain

Self-recognition in animals is demonstrated when individuals pass the mark test. Formerly, it was thought that self-recognition was restricted to humans, great apes, and certain mammals with large brains and highly evolved social cognition. However, 1 study showed that 2 out of 5 magpies (*Pica pica*) passed the mark test, suggesting that magpies have a similar level of cognitive abilities to great apes. The scientific advancement depends on confidence in published science, and this confidence can be reached only after rigorous replication of published studies. Here, we present a close replication of the magpie study but using a larger sample size while following a very similar experimental protocol. Like the previous study, in our experiment, magpies showed both social and self-directed behavior more frequently in front of the mirror versus a control cardboard stimulus. However, during the mark test, self-directed behavior proved more frequent in front of the cardboard than in the mirror. Thus, our replication failed to confirm the previous results. Close replications, while not disproving an earlier study, identify results that should be considered with caution. Therefore, more replication studies and additional experimental work is needed to unambiguously demonstrate that magpies are consistently able to pass the mark test. The existence of compelling evidence of self-recognition in other corvid species is discussed in depth.

Keywords: mark test, *Pica pica*, replication studies, self-directed actions, self-recognition

Mirror self-recognition is a cognitive ability that is so far restricted to humans and a few species with very large brains. The ability to recognize oneself in a mirror is considered by most researchers to be evidence of self-awareness (Gallup, 1998; Rajala, Reininger, Lancaster, & Populin, 2010), which is a form of higher intelligence typical of humans. Although it has been suggested that this point is an open debate because some authors have argued that mirror self-recognition does not necessarily imply self-awareness (Heyes, 1995; Kohda et al., 2019), several articles clearly showing

that it does indicate self-awareness have been published recently (see Bulgarelli et al., 2019; Gallup & Anderson, 2020; Hecht, Mahovetz, Preuss, & Hopkins, 2017; Krachun, Lurz, Mahovetz, & Hopkins, 2019). Mirror self-recognition is often investigated by using the mark test, during which a mark is placed onto a point of the body that cannot be seen without using a mirror (Gallup, 1970). An animal is considered to pass the mark test when it increases the number of actions directed to the mark region in presence of a mirror (experimental condition) relative to number of actions directed to this same area in absence of the mirror or when it was unmarked (control tests).

Since Gallup (1970) designed the mark test, the mirror self-recognition paradigm has been widely used to assess self-recognition in a wide range of species. Individuals may respond by ignoring the mirror image or considering it as a conspecific (de Waal, Dindo, Freeman, & Hall, 2005; Kusayama, Bischof, & Watanabe, 2000; Medina, Taylor, Hunt, & Gray, 2011; Roma et al., 2007; Shaffer & Renner, 2000; Watanabe, 2002) or much more rarely, showing self-directed behavior in front of the mirror, a response that has only been found in just a few species having very large brains (Gallup, 1970; Gallup & Anderson, 2018; Plotnik, de Waal, & Reiss, 2006; Prior, Schwarz, & Güntürkün, 2008; Reiss & Marino, 2001). Until the beginning of the 21st century, only large-brained mammals had passed the mark test, and thus, considering the vast differences between mammals and birds in the organization of their forebrains (Güntürkün, 2012), it was thought that a mammalian neocortex was necessary for self-recognition. However, Prior et al. (2008) found that two out of five magpies (*Pica pica*) were capable of removing the sticker placed under the

 Manuel Soler, Department of Zoology, Faculty of Sciences, and Coevolution Group Unit Associated to the Consejo Superior de Investigaciones Científicas (CSIC), University of Granada; José Manuel Colmenero and  Tomás Pérez-Contreras, Department of Zoology, Faculty of Sciences, University of Granada;  Juan Manuel Peralta-Sánchez, Department of Microbiology, Faculty of Sciences, University of Granada, and Department of Integrative Ecology, Doñana Biological Station, Seville, Spain.

We thank Gordon Gallup Jr. for his constructive and helpful comments. We thank Juan Ontanilla and Laura Arco for their help in the care of the magpies. In addition, Juan Ontanilla for the preparation of Figure 1 and David Nesbitt for improving the English. Manuel Soler and Tomás Pérez-Contreras thank the support provided by Junta de Andalucía group—RNM339. Juan Manuel Peralta-Sánchez was funded by Junta de Andalucía (Proyectos de Excelencia, 2011-RNM-8147).

Correspondence concerning this article should be addressed to Manuel Soler, Departamento de Zoología, Facultad de Ciencias, Universidad de Granada, E-18071 Granada, Spain. E-mail: msoler@ugr.es

beak, outside the magpies' visual field, by scratching with their foot in mirror-present sessions but not during the nonreflective plate sessions.

The biological and cognitive implications of this experimental study are far-reaching. The conclusion that magpies have the self-recognition ability, a capacity that has not been documented in most primate species (de Waal et al., 2005), signifies, as Prior et al. (2008) emphasized, that magpies are at a similar level of cognitive abilities than chimpanzees (*Pan troglodytes*). This implies that mirror self-recognition has evolved independently in the magpie and great apes, which diverged 300 million years ago. In addition, this study provided evidence showing that neocortex is not a requirement for self-recognition (Güntürkün, 2012; Prior et al., 2008). This finding was not surprising because corvids have allometrically large brains (Jerison, 1969) and possess a repertoire of complex cognitive abilities that are on par with apes (Emery & Clayton, 2004), which has been suggested as evidence of convergence between the mental abilities of corvids and primates (Emery & Clayton, 2004; Güntürkün, 2012).

The conclusion that individuals of one species recognize themselves in the mirror or are able to pass the mark test is not free from controversy. Absence of self-directed behavior in the mark test may lead to false-negative results (Uchino & Watanabe, 2014), but false positives (i.e., considering that one individual has passed the mark test when the mark is touched or removed by chance) may also occur (Soler, Pérez-Contreras, & Peralta-Sánchez, 2014), as is frequent in general in scientific studies (Parker et al., 2016). In fact, in a recent review, it has been stated that "despite claims to the contrary no dogs, elephants, dolphins, magpies, horses, manta rays, squid, or ants have shown compelling, reproducible evidence for self-recognition" (Gallup & Anderson, 2018, p. 16). Therefore, replication of empirical evidence is crucial in general and particularly in the self-recognition paradigm; the lack of replicate studies contributes to maintain previous errors, which greatly retard the advancement of biological and evolutionary science (Palmer, 2000).

As scientific advancement generates new knowledge based on previous findings, the validity of the process depends on confidence in published science, but sometimes original measurements can be the result of a chance convergence of unknown factors. In fact, false positives occur in at least 5% of all published findings (Parker et al., 2016), and any single result should be viewed only as preliminary and suggestive (Parker & Nakagawa, 2017). Thus, confidence in published results can be reached only after rigorous replication of previously published studies (Forstmeier, Wagenmakers, & Parker, 2017; Parker & Nakagawa, 2017; Simons, 2014). Replication is an essential part of cumulative science (Forstmeier et al., 2017; Kelly, 2006; Parker & Nakagawa, 2017) and promotes scientific advancement by both assessing the validity of previous findings and testing their generality (Nakagawa & Parker, 2015). Rigorous and well-conducted replications constitute a robust estimate of the size and reliability of the original finding (Forstmeier et al., 2017; Simons, 2014).

The carefully designed and well-controlled experiment made by Prior et al. (2008) in magpies was quasi-replicated in jackdaws (*Corvus monedula*) with the aim of determining whether this large-brained corvid with complex social behavior and sophisticated cognitive abilities (Lorenz, 1982; Röell, 1978) was also capable of passing the mark test (Soler et al., 2014). This repetition

of studies using different species or systems is frequently used in behavioral and cognitive sciences and are known as quasi-replications (Palmer, 2000). The study by Soler et al. (2014) failed to provide evidence of self-recognition in jackdaws (as previously occurred with other corvid species; see Kusayama et al., 2000; Medina et al., 2011). Jackdaws consistently showed interest in the mirror, spending significantly more time viewing and inspecting it than the cardboard, and self-contingent behavior (comparison between one's own movements and those of the reflection) started as soon as the mirror was placed (Soler et al., 2014), whereas magpies showed self-contingent behavior only after 150 min of mirror exposure (Prior et al., 2008). Intensive mirror-image exploration is an important requirement to pass the mark test (Heschl & Fuchs-bichler, 2009). Indeed magpies that showed a high preference for the mirror, later performed mark-directed actions (Prior et al., 2008). Although jackdaws presented one of the highest levels of interest in the mirror so far reported, including great apes that usually do pass the mark test (Gallup, 1970; Povinelli, Rulf, Landau, & Bierschwale, 1993), they performed mark-directed actions and removed stickers in front of the cardboard as well as in front of the mirror (Soler et al., 2014). Jackdaws' findings, together with the fact that magpies also performed mark-directed actions under nonmirror conditions (Prior et al., 2008) suggest that the most parsimonious explanation for the mark-directed actions reported in both species is by tactile cues provided by the marks, rather than mirror use (Soler et al., 2014).

Quasi-replications, although useful, are often insufficient for reliably determining the true reason for the difference in results because quasi-replicates differ in many aspects (Forstmeier et al., 2017; Nakagawa & Parker, 2015). However, quasi-replications are important because they demonstrate that the results are robust and generalizable, two points that are of crucial importance in science. Exact replication is theoretically the best type of replication. However, exact replications are impossible in biological sciences, and thus, attempts of exact replications may remain, at best, "close" replications (Nakagawa & Parker, 2015). When replications include limited procedural or methodological differences they are known as "partial" replications (Nakagawa & Parker, 2015). Here, we present a close replication of the study of Prior et al. (2008) by using eight magpies captured in the wild and following almost exactly the five successive experimental stages and performing the same statistical analyses as in Prior et al. (2008).

Materials and Method

Study Subjects and Housing

Eight magpies of unknown age and sex were captured in the wild during January, 2013. Using wild birds instead of hand reared birds as Prior et al. (2008) could add some uncontrolled factors to the study. For instance, this does not allow us to control for previous experiences with reflective surfaces, and in birds not used to human handling, the experimental procedure could be expected to be more stressful than for hand-raised birds. To minimize this possibility, the subjects were maintained in captivity for 1 year before the mirror experiment to keep them accustomed to captivity conditions. We additionally think that this difference is not important because in the jackdaw study (Soler et al., 2014) most of the individuals were hand raised, and the results were very similar to

the ones found in the present study. Magpies were identified with a unique combination of colored leg bands for individual identification, and they are named in the text and in tables according to their colored leg bands. They were not used in any other experiment. The subjects were maintained in two outdoor aviaries of 50 m³ each, interconnected by a hole of about 25 cm in diameter. These aviaries were located in the Hoya de Guadix (southern Spain, a high-altitude plateau, approx. 1,000 m a.s.l., near Hernán Valle, 60 km from Granada). The magpies were provided with bread, apple, lettuce, fodder for dog puppies, and had ad libitum access to water.

Experimental Conditions

During February and March, 2014, the experimental sessions were conducted in a box made of particle board and wire mesh (160 × 100 × 80 cm; length × height × width) without perches (see Figure 1 in Soler et al., 2014, for a detailed description of the box used in the experimental sessions), keeping the experimental subjects on the ground, at the same level as the mirror. The experimental box could easily be adapted to the necessities of each experimental test by making it smaller (fourth experimental stage) or separating it with two sheets of particle board into two identical compartments of 70 × 100 × 80 cm, divided by a corridor 20 cm wide (compartmentalized experimental box; third experimental stage). In this case, the entrance to one compartment was closer than the entrance to the other, and when the bird was released one entrance was closer than the other entrance. Therefore, this experimental box allowed the magpies to move freely between compartments, but prevented them from seeing from one compartment into the other. The front of the experimental box consisted of wire mesh, which allowed direct observation of both compartments and the corridor. The mirror, 60 × 50 cm (height × width), was placed directly on the ground and in a vertical position because in this way it is more effective at eliciting “mirror-image”-directed responses than when placed in a horizontal position (Kusayama et al., 2000; Pepperberg, Garcia, Jackson, & Marconi, 1995). The

mirror was thoroughly cleaned before each experimental session, and such cleaning did not provide olfactory cues.

During experimental tests, all magpies were moved to one of the aviaries, and the experimental box was placed in the other, closing the hole that connected both aviaries. Then, each individual was captured, placed in an ornithological cloth bag, taken to the other aviary and released in the experimental box through the entrance of the corridor.

Experimental Design

We generally followed the experimental procedure as described by Prior et al. (2008) consisting of five successive experimental stages. In Stage 1, each bird was exposed for 30 min to a baseline session in which a vertical mirror-sized gray nonreflective cardboard was set in the large experimental box. The cardboard was set 50 cm away from the wall, enabling the subjects to move freely around it. In Stage 2, each bird was exposed to the vertical mirror in three subsequent daily sessions of 50 min each, in the large experimental box, in which, again, the mirror was set 50 cm away from the wall. During Stage 2, Prior et al. (2008) performed five 30-min test sessions, as a result, we reduced the number of sessions while increasing the duration of each session in a way that the total exposure time in our study was identical to theirs. By reducing the number of sessions, we minimized the stress of the experimental magpies. As far as we know, the effect of the length of sessions in mirror exposure on habituation has never been studied, but we think there is no reason to expect that decreasing the number of sessions while maintaining the total exposure time might affect the results.

Stage 3 was devoted to quantifying preference for the mirror and mirror-induced behavior. During this stage, each magpie was exposed to the compartmentalized experimental box in six (five in Prior et al., 2008), 20-min sessions on two different days. Birds were submitted to the next 20-min session only after all the other subjects had been tested (i.e., about 3 h after the previous 20-min session). In every session, each subject was released in the corridor between compartments and could move freely from one to the other. The position of the mirror and the cardboard was switched from one compartment to the other on the second day.

Stage 4 constitutes the mark test itself. During this stage, each magpie was involved in eight mark-test sessions with the mirror or the cardboard placed on the wall. These sessions included all four conditions: (a) mirror and colored (yellow or red) sticker, (b) mirror and black sticker, (c) cardboard and colored (yellow or red) sticker, and (d) cardboard with black sticker. The weight of the stickers was 2,375 and 3,960 μg for 6-mm and 10-mm stickers, respectively. The mirror and the cardboard were of the same size and were placed in the same position vertically. Each session lasted 20 min, and each condition was replicated once. We allowed at least 3 h between two consecutive sessions, and we randomized conditions and the order of presentation. These experimental sessions were conducted using only one of the compartments (including the corridor) of our experimental box (please refer the earlier text).

Stage 4 was repeated twice: In the first, smaller (6-mm) stickers were used, and in the second, larger (10-mm) stickers were used in comparison with the 8-mm stickers used by Prior et al. (2008; see Figure 1). This use of smaller and larger stickers would allow



Figure 1. Magpies with the large (10 mm, above) and small (6 mm, below) stickers. See the online article for the color version of this figure.

exploring further the methodology of mark tests because it makes possible to test the prediction that individual responses to the mark should be higher with the larger than with the smaller marks. We fixed the self-adhesive stickers onto the throat region while keeping the bird's head inside the bag. Thus, the bird was prevented from seeing anything during the fixing procedure. The stickers fixed onto the throat region (see Figure 1) were outside the magpies' visual field (Prior et al., 2008). The action of pressing the sticker on the throat feathers was repeated three to five times on the breast and the wing.

Scoring of Behavior

In all experimental sessions, the behavior of the birds was directly observed through the windows of a car parked about 30 m from the aviary. We are confident that the presence of the observer inside the car did not affect the behavior of the magpies in the experimental box, given that, in the second aviary, magpies maintained normal behavior. In addition, we videotaped all experimental sessions using a Panasonic HDC-SD40 camera. We scored the birds' behavior in a way similar to that of Prior et al. (2008). From the videotape recordings, we quantified the following: (a) time with a view of the mirror or the cardboard; (b) time of close inspection of the mirror or cardboard (i.e., subject stared directly at the mirror or cardboard at close range); (c) frequency of pecks directed to the mirror or cardboard; (d) frequency of looking behind the mirror or cardboard; (e) frequency of self-contingent behavior (see the following text); (f) frequency of social behavior (i.e., antagonistic or submissive displays); (g) frequency of self-contingent behavior (directed toward sticker-less body parts); and (h) frequency of mark-directed behavior (specifically, directed toward the throat region). Following Prior et al. (2008), we considered the bird's jumping and/or flying toward the mirror as social behavior; however, we consider such movement to be, at least sometimes, only an attempt to go through the mirror as also occurred in jackdaws (Soler et al., 2014).

The behavior was considered an indication of self-contingency when the subject moved in front of the mirror in a systematic way, as if assessing the relationship between the mirror image and its own movements. Prior et al. (2008) considered magpie movements

of the head or the whole body back and forth and left and right in front of the mirror to be contingent behavior. Two of the authors, José Manuel Colmenero and Juan Manuel Peralta-Sánchez, independently scored the behavior of the magpies based on the video recordings of 20 bird/sessions (15.63%) randomly chosen, and their scores were highly correlated both when considering behavioral variables ($N = 116$, $r_s = 0.94$, $p < .001$) and when considering variables related to quantification of time ($N = 58$, $r_s = 0.96$, $p < .001$). Wilcoxon's matched-pairs tests were performed in Statistica 10.0 and Fisher-exact tests (the same used by Prior et al., 2008) in R environment (R Core Team, 2015).

Ethical Note

The research was conducted according to national (Real Decreto 1201/2005, de 10 de Octubre) and regional guidelines and was approved by the "Comisión de Ética en Investigación de la Universidad de Granada." All necessary permits, including that for confinement of magpies, were obtained from the "Consejería de Medio Ambiente de la Junta de Andalucía," Spain. All efforts were made to minimize suffering, and no bird showed symptoms of stress or died during this study. The magpies' well-being was followed by monitoring their physical condition when providing the food every 2–3 days and before performing the present experiment: All birds were captured, measured, and examined in detail. After this study, magpies were released.

Results

Behavior in Front of Mirror or Cardboard

During Stages 1 and 2, magpies spent more time in front of the mirror and looked at it more frequently than at the cardboard (see Table 1). Self-contingent and social behavior also proved more frequent in front of the mirror, being practically nonexistent in the cardboard stage (see Table 1). For the variables related to mirror/cardboard inspection, there were no significant differences in the frequency of looking behind the mirror/cardboard and the frequency of pecks (see Table 1). In Stage 3, where the magpies could choose between the two compartments (one with mirror and an-

Table 1

Magpies' Behavioral Data for Experimental Stages 1 (30-Min Baseline Session With Cardboard) and 2 (150-Min Mirror Exposure Sessions)

Individuals	Time with view of (s/hr)		Close inspection of (s/hr)		Frequency of pecks to		Frequency of looks behind		Frequency of social behaviors		Frequency of self-contingent behavior	
	Mirror	Cardboard	Mirror	Cardboard	Mirror	Cardboard	Mirror	Cardboard	Mirror	Cardboard	Mirror	Cardboard
Double yellow	610.40	20	68.80	2	2.40	0	8.80	2	1.60	2	3.20	0
Right orange	2,698.40	3,228	91.20	26	2.80	0	12.40	24	9.60	0	4.40	0
Right green	3,186.40	214	312.40	14	2.00	4	3.20	14	0.40	0	4.00	0
Right blue	1,170.40	0	87.60	0	1.20	0	25.20	0	7.20	0	4.00	0
Red	1,350.00	470	48.00	22	0.00	0	0.40	2	9.60	0	0.80	0
Double red	1,196.00	160	132.40	10	810.80	0	49.20	4	6.80	0	8.40	0
Double blue	2,410.00	106	20.00	6	20.80	0	0.00	0	2.40	0	1.60	0
Double green	814.80	0	814.40	0	1.60	0	0.40	0	2.00	0	1.20	0
Wilcoxon matched pairs test	$T = 1$; $p = .017$		$T = 0$; $p = .012$		$T = 3$; $p = .063$		$T = 11$; $p = .612$		$T = 1$; $p = .017$		$T = 0$; $p = .012$	

Note. Significant p values are shown in bold.

other with cardboard), the magpies spent significantly more time looking at the mirror than at the cardboard, but there were no differences with respect to the time they spent in the two compartments, either in the number of pecks or the behavior of looking and then walking behind the mirror (see Table 2). The self-contingent and social behavior was not detected in relation to the cardboard, and only the self-contingent behavior significantly differed between cardboard and mirror compartment (see Table 2).

At Stage 3, magpies showed no preference for the first room choice. However, they preferred to enter the room closer to the entrance without taking into account whether the mirror or the cardboard was in place. In 25 of the 32 tests performed (78.12%), the magpies chose the compartment whose entrance was closer to the entrance door, whereas on five occasions (15.62%) they entered the room farther away, and on only two occasions (6.25%) did the magpie stay in the central corridor during the complete duration of the session (20 min, Table 3). In both cases, it was the double yellow magpie.

Mark Test

In Stage 4 (mark test), the magpies showed little motivation during the session with small stickers (two actions), and both actions were made by the magpie right blue; one act was a mark directed and the other self-directed (beak directed at one of the wings), in both cases facing the mirror with the black sticker (see Table 3).

Magpies were more active with the large stickers (18 actions), although none of these acts was performed by right blue. In this session, two magpies acted directly toward the mark, and also both acted toward other parts of the body (wings, legs, and abdomen). Magpie red performed two actions in front of the mirror with a colored sticker, one toward the mark and the other toward the abdomen. Magpie double yellow also performed mark-directed and self-directed actions, but all of them ($n = 16$) facing the cardboard (Fisher-exact tests, $p = .604$). Of the 16 actions, 12 were against the colored stickers, and the remaining four to black stickers (see Table 3).

Discussion

Prior et al. (2008) found that magpies showed both social and self-directed behavior more frequently in front of the mirror than

in the cardboard situation, and that mark-directed behavior occurred during the mark test, trying to remove the sticker in front of the mirror and removing the sticker in two cases. In our study, the magpies' behavior was also consistent regarding the presence of the mirror or cardboard. During Stages 1 and 2, both social and self-directed behavior proved more frequent in front of the mirror than in the cardboard situation (see Table 1), and during experimental Stage 3, self-directed behavior was also more frequent in front of the mirror than with the cardboard (see Table 2). However, during the mark test (Stage 4), six out of eight mark-directed actions were performed during the cardboard situation. These actions were accounted for by four magpies that responded to cardboard or mirror, where only two magpies performed mark-directed actions (right blue and red) in front of the mirror. Thus, our replication fails to confirm the original results by Prior et al. (2008).

False-Negative or False-Positive Findings?

However, these findings do not mean that our results invalidate the conclusions of Prior et al. (2008). Close replications that produce the same results increase confidence in the original study, but unsuccessful close replications, though decreasing confidence, do not prove that an earlier study is wrong but rather only identify results that should be considered with caution (Kelly, 2006; Nakagawa & Parker, 2015). Failure to replicate the results of Prior et al. (2008) could be for two main reasons. First, behavioral and cognitive studies cannot be perfectly replicated because both behavior and cognitive abilities are highly complex and variable among individuals. The context-dependence explanation, which is frequently used when a replication fails to confirm the original results (Forstmeier et al., 2017), applies in this case. That is, the magpies used in this study and those used by Prior et al. (2008) differ in many ways. For instance, the magpies in the study by Prior et al. (2008), contrary to our magpies, were hand raised and were previously used in other experimental studies. In addition, conditions of confinement likely differed. Second, self-recognition is not a ubiquitous feature shared by all individuals in one species, given that in mark-test studies, usually only a small proportion of individuals pass the test (Povinelli et al., 1993; Prior et al., 2008). Thus, it could be possible by chance that there would be two self-recognizing magpies in the sample of five in Prior et al. (2008)

Table 2
Magpies' Behavioral Data for Experimental Stage 3 (Choice Between Mirror and Cardboard Compartments)

Individuals	Time with view of (s/hr)		Close inspection of (s/hr)		Frequency of pecks to		Frequency of looks behind		Frequency of social behaviors		Frequency of self-contingent behavior	
	Mirror	Cardboard	Mirror	Cardboard	Mirror	Cardboard	Mirror	Cardboard	Mirror	Cardboard	Mirror	Cardboard
Double yellow	900	900	116.25	26.25	0	0.75	0	0	0	0	0	0
Right orange	2,553.75	207.75	77.25	15.00	1.50	0	0	0	0	0	0	0
Right green	1,781.25	1,812.00	504.00	299.25	0	0.75	0	0	0	0	0	0
Right blue	1,926.75	1,341.75	249.75	169.50	0	0	0.38	0	0	0	84.75	0
Red	1,713.75	1,635.75	141	110.25	0	0	0	3.75	16.5	0	14.25	0
Double red	2,214	1,059	33.75	13.5	0	0	0	0	0	0	4.5	0
Double blue	2,070.75	1,581	108.75	13.5	0	0	0	0	0	0	1.5	0
Double green	968.25	2,427	30	33	1.5	0	0	3.75	1.5	0	4.5	0
Wilcoxon matched pairs test	$T = 7, p = .237$		$T = 1, p = .017$		$T = 3, p = .465$		$T = 1, p = .285$		$T = 0, p = .180$		$T = 0, p = .043$	

Note. Significant p values are shown in bold.

Table 3
Maggies' Frequencies of Self-Directed Behaviors in Experimental Stage 4 (Mark Test)

Individuals	Small stickers				Large stickers			
	Cardboard/Color	Cardboard/Black	Mirror/Color	Mirror/Black	Cardboard/Color	Cardboard/Black	Mirror/Color	Mirror/Black
Double yellow	0/0	0/0	0/0	0/0	4/8	2/2	0/0	0/0
Right orange	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Right green	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Right blue	0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0
Red	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/0
Double red	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Double blue	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Double green	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

Note. Double yellow magpie did not show significant differences between cardboard/color condition and other conditions (Fisher-exact tests, $p > .604$). In each case, the first number gives the number of mark-directed actions, and the second refers to the number of self-directed actions toward other parts of the body.

and none in our sample of eight, which would lead to false-negative results in our study. However, the opposite explanation, that is, false-positive results in Prior et al. (2008) study, seems at least equally likely owing to the considerable difference between the results found in Prior et al. (2008) and the present study, but also because false-positive results are more frequent in science and increase with small sample sizes (Forstmeier et al., 2017). Therefore, we may conclude that our results suggest that the most parsimonious explanation for the differences reported in Prior et al. (2008) and this study is that their results were the product of chance.

This statement is supported by three additional pieces of evidence. First, one of the main criticisms made against mirror self-recognition studies is that an individual could make self-directed actions without using its reflection in the mirror (Povinelli et al., 1993; Bard, Todd, Bernier, Love, & Leavens, 2006; Suddendorf & Collier-Baker, 2009). This seems to be the case with magpies, given that in the Prior et al. (2008) study, they performed mark-directed actions also in nonmirror conditions, and, in our study, mark-directed actions showed a trend to be more frequent in the nonmirror experimental situation. This suggests that magpies are able to detect the sticker independently of the sense of vision, likely using the sense of touch (Soler et al., 2014). Second, methodological issues have often been suggested as a potential source of erroneous results in mark test experiments (Bard et al., 2006; De Veer & van den Bos, 1999; Medina et al., 2011; Suddendorf & Collier-Baker, 2009). Two of these issues may be important in magpie studies: the use of stickers as an appropriate replacement for the paint marks used in mammalian mirror studies, and the virtual impossibility of producing a true sham mark control because of the iridescence of magpie feathers. And third, as pointed out by Gallup and Anderson (2018), the two magpies presenting mark-directed behavior in the Prior et al. (2008) study persisted repeatedly in the responses toward their reflection for extended periods of time. This lack of habituation to the mirror is consistent with findings in species not able to pass the mark test, such as monkeys and parrots (Gallup & Suarez, 1991; Pepperberg et al., 1995). If magpies have the ability of self-recognition it would be expected for them to habituate to the mark, as occurs in chimpanzees (*Pan troglodytes*): When they realize that it is their behavior that they are observing in the mirror, their interest rapidly

diminishes, whereas their self-contingent and mirror-directed exploratory behavior increases (Gallup & Anderson, 2018).

Is There Evidence of Self-Recognition in Other Corvid Species?

Corvids, a group of birds that are known to have allometrically large brains (Jerison, 1969), have been found to possess a repertoire of complex cognitive abilities that are on par with those of primates (Emery & Clayton, 2004; Marzluff & Angell, 2012). For this reason, the results of Prior et al. (2008) showing self-recognition in magpies were not surprising. But, what happens in other corvid species? Mirror-induced self-directed behavior has been studied in seven other corvid species. Jungle crows (*Corvus macrorhynchos*) and jackdaws showed a high interest in the mirror, but their social and aggressive behavior indicated that both species perceive their mirror reflection as a conspecific, not as their own image (Kusayama et al., 2000; Soler et al., 2014). These results are also confirmed by the fact that they did not habituate to their mirror reflections, that is, their social responses were maintained over time.

New Caledonian crows (*Corvus moneduloides*), a corvid species with complex cognition and a higher capacity for solving problems, behavioral innovation, and tool manufacture (Hunt & Gray, 2003; Taylor, Hunt, Medina, & Gray, 2009, 2010), have also been found to engage in aggressive behavior when confronted with their mirror reflection without habituating to their mirror image (Medina et al., 2011).

Recently, Clary, Stow, Vernouillet, and Kelly (2019) have found that California scrub jays (*Aphelocoma californica*) showed no evidence of attempting to remove the mark in the presence of the mirror, and Vanhooland, Bugnyar, and Massen (2019) have shown that both the carrion and hooded crows (*Corvus corone* ssp.), as other corvid species, show a high interest in the mirror, although none of the individuals showed significant mark-directed behaviors during the mark test.

In contrast of these five species in which no evidence of mark-directed behavior has been found, other authors claim the self-recognition in two other corvid species during the mark test: the Indian house crow (*Corvus splendens*) and the Clark's nutcracker (*Nucifraga columbiana*). Buniyaadi, Taufique, and Kumar (2019)

have shown that Indian house crows have the ability of self-recognition, being the proportion of crows that responded to the mark in mirror conditions greater than in other corvids studied previously.

Clark's nutcracker deserves special mention because the authors stated: "this is the first study to show a corvid self-recognizing during the mark test that cannot be accounted for by the use of tactile cues, providing evidence of convergence between the mental abilities of corvids and primates and that a mammalian neocortex is not necessary for self-recognition" (Clary & Kelly, 2016, p. 9). This strong assertion that their results during the mark test are not accounted for by the use of tactile cues is not supported. The fact that nutcrackers do not ruffle their feathers significantly more often after mark application does not unambiguously demonstrate that they do not use tactile cues to detect the mark. Support for that conclusion would require an experiment specifically designed to test that hypothesis. In any case, the relevant question is: Why not? The skin of birds has many receptors sensitive to pressure and touch (Birkhead, 2012).

We suggest that the results obtained in the mark test experiment in both studies are unwarranted because their results are not clear enough and many questions arise, some regarding the Clark's nutcracker study, and some regarding both studies. With respect to Clark's nutcrackers:

(1) Why did 40% of the birds perform the actions directed to the mark mainly during the barrier conditions (Table 1 in Clary & Kelly, 2016)?

(2) The conclusion of Clary and Kelly (2016) that self-recognition exists in Clark's nutcrackers is based on more mark-directed actions in a blurry mirror, but not under regular mirror conditions. Why would nutcrackers make more mark-directed actions with a blurry mirror than under regular mirror conditions? The idea that viewing a mirror provides two types of information: identity and contingent motion information is appealing, but it seems to be an a posteriori explanation because it could not be predicted a priori, and, furthermore, it does not convincingly explain the results. It is true that a blurry mirror provides only contingent motion information, preventing distinguishing a clear identity. However, a regular mirror provides both contingent motion information and clear identity information. Why would the lack of identity information favor self-recognition ability? Why should favor a blurry mirror, which is a different dimension of visual experience under less favorable conditions, more mark-directed actions than a regular mirror conditions? In addition, the crucial point for nutcrackers in the mark test is to see the mark, and, as showed in Figure 2c in Clary and Kelly (2016), we can barely make out the red mark in the blurry mirror.

(3) How can it be concluded that Clark's nutcrackers are capable of recognizing themselves in a mirror when there are no significant differences (Figure 4a in Clary & Kelly, 2016; before removing from the analyses those individuals whose behavior does not agree with predictions) between the number of mark-directed actions performed in the experimental (blurry or regular mirror) and control (barrier) conditions?

Regarding both studies:

(1) In the Indian house crow, most of the responses during the mark-mirror conditions (30 out of 34) are performed by Individuals 5 and 6, which also performed most of the head shakes, plumage ruffling, and preening behaviors (30 out of 46, 33 out of

42, and 85 out of 107, respectively). In the Clark's nutcrackers, a correlation existed ($N = 60$; $r = .153$; $p = .002$) between the number of actions directed to the mark and number of actions directed to other parts of the body (presented in Table 1; Clary & Kelly, 2016). Do both results indicate that mark-directed actions can be explained simply by a general increase of behavioral activity?

(2) Were Indian house crow and Clark's nutcrackers capable of removing the sticker in any case? Apparently not, as it has not been reported. Some jackdaws and magpies did (Prior et al., 2008; Soler et al., 2014). An individual with self-recognition capacity would be expected to be successful in removing the mark. In fact, Prior et al. (2008) repeated the mark test for Gerti, the magpie who showed the most clear and consistent self-directed behavior, and it removed the mark in the four additional tests of 5-min periods (including the control black mark).

(3) And, finally, why did Indian house crows and Clark's nutcrackers not show habituation to the mirror as occurs in species able to pass the mark test? (see the earlier text).

Conclusions

No conclusive evidence of self-recognition has so far been reported in any corvid species. We do not mean that corvids are incapable of mirror self-recognition, but rather we advise caution while recommending that more replication studies and additional experimental work is needed to unambiguously demonstrate that one corvid species is able to consistently pass the mark test. The previously described case of Gerti in the study of Prior et al. (2008) that removed the mark in all additional tests is an intriguing and challenging matter of research. And, with respect to Indian house crows and Clark's nutcrackers, the mark-test experiment needs to be replicated using improved methodology that resolves the methodological problems related to the use of stickers and the impossibility of producing a true sham mark control; for instance, painting the feathers with typing correction fluid, which allow the feathers to be separated when dried, which would prevent the mark being detected by the use of tactile cues (Soler et al., 2014). Thus, to make the assertion that one corvid species has the capacity of self-recognition, which would have evolved independently of great apes about 300 million years ago, implying that the neocortex is not a prerequisite for mirror self-recognition as previously believed (Clary & Kelly, 2016; Prior et al., 2008), is far from straightforward.

References

- Bard, K., Todd, B. K., Bernier, C., Love, J., & Leavens, D. A. (2006). Self-awareness in human and chimpanzee infants: What is measured and what is meant by the mark and mirror test. *Infancy*, 9, 191–219. http://dx.doi.org/10.1207/s15327078in0902_6
- Birkhead, T. (2012). *Bird sense: What it's like to be a bird*. London, UK: Bloomsbury.
- Bulgarelli, C., Blasi, A., de Klerk, C. C. J. M., Richards, J. E., Hamilton, A., & Southgate, V. (2019). Fronto-temporoparietal connectivity and self-awareness in 18-month-olds: A resting state fNIRS study. *Developmental Cognitive Neuroscience*, 38, 100676. <http://dx.doi.org/10.1016/j.dcn.2019.100676>
- Buniyaadi, A., Taufique, S. K. T., & Kumar, V. (2019). Self-recognition in corvids: Evidence from the mirror-mark test in Indian house crows

- (*Corvus splendens*). *Journal of Ornithology*, *161*, 341–350. <http://dx.doi.org/10.1007/s10336-019-01730-2>
- Clary, D., & Kelly, D. M. (2016). Graded mirror self-recognition by Clark's Nutcrackers. *Scientific Reports*, *6*, 36459. <http://dx.doi.org/10.1038/srep36459>
- Clary, D., Stow, M. K., Vernouillet, A., & Kelly, D. M. (2019). Mirror-mediated responses of California scrub jays (*Aphelocoma californica*) during a caching task and the mark test. *Ethology*, *126*, 140–152. <http://dx.doi.org/10.1111/eth.12954>
- De Veer, M. W., & van den Bos, R. (1999). A critical review of methodology and interpretation of mirror self-recognition research in nonhuman primates. *Animal Behaviour*, *58*, 459–468. <http://dx.doi.org/10.1006/anbe.1999.1166>
- de Waal, F. B. M., Dindo, M., Freeman, C. A., & Hall, M. J. (2005). The monkey in the mirror: Hardly a stranger. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 11140–11147. <http://dx.doi.org/10.1073/pnas.0503935102>
- Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, *306*, 1903–1907. <http://dx.doi.org/10.1126/science.1098410>
- Forstmeier, W., Wagenmakers, E.-J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings - a practical guide. *Biological Reviews of the Cambridge Philosophical Society*, *92*, 1941–1968. <http://dx.doi.org/10.1111/brv.12315>
- Gallup, G. G., Jr. (1970). Chimpanzees: Self-recognition. *Science*, *167*, 86–87. <http://dx.doi.org/10.1126/science.167.3914.86>
- Gallup, G. G., Jr. (1998). Self-awareness and the evolution of social intelligence. *Behavioural Processes*, *42*, 239–247. [http://dx.doi.org/10.1016/S0376-6357\(97\)00079-X](http://dx.doi.org/10.1016/S0376-6357(97)00079-X)
- Gallup, G. G., Jr., & Anderson, J. R. (2018). The “olfactory mirror” and other recent attempts to demonstrate self-recognition in non-primate species. *Behavioural Processes*, *148*, 16–19. <http://dx.doi.org/10.1016/j.beproc.2017.12.010>
- Gallup, G. G., Jr., & Anderson, J. R. (2020). Self-awareness in animals: Where do we stand 50 years later? Lessons from cleaner wrasse and other species. *Psychology of Consciousness: Theory, Research, and Practice*, *7*, 46–58. <http://dx.doi.org/10.1037/cns0000206>
- Gallup, G. G., Jr., & Suarez, S. D. (1991). Social responding to mirrors in rhesus monkeys (*Macaca mulatta*): Effects of temporary mirror removal. *Journal of Comparative Psychology*, *105*, 376–379. <http://dx.doi.org/10.1037/0735-7036.105.4.376>
- Güntürkün, O. (2012). The convergent evolution of neural substrates for cognition. *Psychological Research*, *76*, 212–219. <http://dx.doi.org/10.1007/s00426-011-0377-9>
- Hecht, E. E., Mahovetz, L. M., Preuss, T. M., & Hopkins, W. D. (2017). A neuroanatomical predictor of mirror self-recognition in chimpanzees. *Social Cognitive and Affective Neuroscience*, *12*, 37–48. <http://dx.doi.org/10.1093/scan/nsw159>
- Heschl, A., & Fuchsichler, C. (2009). Siamangs (*Hylobates syndactylus*) recognize their mirror image. *International Journal of Comparative Psychology*, *22*, 221–233.
- Heyes, C. M. (1995). Self-recognition in primates: Further reflections create a hall of mirrors. *Animal Behaviour*, *50*, 1533–1542. [http://dx.doi.org/10.1016/0003-3472\(95\)80009-3](http://dx.doi.org/10.1016/0003-3472(95)80009-3)
- Hunt, G. R., & Gray, R. D. (2003). Diversification and cumulative evolution in New Caledonian crow tool manufacture. *Proceedings Biological Sciences*, *270*, 867–874. <http://dx.doi.org/10.1098/rspb.2002.2302>
- Jerison, H. J. (1969). Brain evolution and dinosaur brains. *American Naturalist*, *103*, 575–588. <http://dx.doi.org/10.1086/282627>
- Kelly, C. D. (2006). Replicating empirical research in behavioral ecology: How and why it should be done but rarely ever is. *The Quarterly Review of Biology*, *81*, 221–236. <http://dx.doi.org/10.1086/506236>
- Kohda, M., Hotta, T., Takeyama, T., Awata, S., Tanaka, H., Asai, J. Y., & Jordan, A. L. (2019). If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals? *PLoS Biology*, *17*, e3000021. <http://dx.doi.org/10.1371/journal.pbio.3000021>
- Krachun, C., Lurz, R., Mahovetz, L. M., & Hopkins, W. D. (2019). Mirror self-recognition and its relationship to social cognition in chimpanzees. *Animal Cognition*, *22*, 1171–1183. <http://dx.doi.org/10.1007/s10071-019-01309-7>
- Kusayama, T., Bischof, H.-J., & Watanabe, S. (2000). Responses to mirror-image stimulation in jungle crows (*Corvus macrorhynchos*). *Animal Cognition*, *3*, 61–64. <http://dx.doi.org/10.1007/s100710050051>
- Lorenz, K. (1982). *Hablaba con las bestias, los peces y los pájaros* [Talking with beasts, fishes and birds]. Barcelona, Spain: Labor S. A.
- Marzluff, J., & Angell, T. (2012). *Gifts of the crow: How perception, emotion, and thought allow smart birds to behave like humans*. New York, NY: Atria.
- Medina, F. S., Taylor, A. H., Hunt, G. R., & Gray, R. D. (2011). New Caledonian crows' responses to mirrors. *Animal Behaviour*, *82*, 981–993. <http://dx.doi.org/10.1016/j.anbehav.2011.07.033>
- Nakagawa, S., & Parker, T. H. (2015). Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, *13*, 88. <http://dx.doi.org/10.1186/s12915-015-0196-3>
- Palmer, A. R. (2000). Quasi-replication and the contract of error: Lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annual Review of Ecology and Systematics*, *31*, 441–480. <http://dx.doi.org/10.1146/annurev.ecolsys.31.1.441>
- Parker, T. H., Forstmeier, W., Koricheva, J., Fidler, F., Hadfield, J. D., Chee, Y. E., . . . Nakagawa, S. (2016). Transparency in ecology and evolution: Real problems, real solutions. *Trends in Ecology and Evolution*, *31*, 711–719. <http://dx.doi.org/10.1016/j.tree.2016.07.002>
- Parker, T. H., & Nakagawa, S. (2017). Practical models for publishing replications in behavioral ecology: A comment on Ihle et al. *Behavioral Ecology*, *28*, 355–357. <http://dx.doi.org/10.1093/beheco/arw191>
- Pepperberg, I. M., Garcia, S. E., Jackson, E. C., & Marconi, S. (1995). Mirror use by African gray parrots (*Psittacus erithacus*). *Journal of Comparative Psychology*, *109*, 182–195. <http://dx.doi.org/10.1037/0735-7036.109.2.182>
- Plotnik, J. M., de Waal, F. B. M., & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 17053–17057. <http://dx.doi.org/10.1073/pnas.0608062103>
- Povinelli, D. J., Rulf, A. B., Landau, K. R., & Bierschwale, D. T. (1993). Self-recognition in chimpanzees (*Pan troglodytes*): Distribution, ontogeny, and patterns of emergence. *Journal of Comparative Psychology*, *107*, 347–372. <http://dx.doi.org/10.1037/0735-7036.107.4.347>
- Prior, H., Schwarz, A., & Güntürkün, O. (2008). Mirror-induced behavior in the magpie (*Pica pica*): Evidence of self-recognition. *PLoS Biology*, *6*, e202. <http://dx.doi.org/10.1371/journal.pbio.0060202>
- Rajala, A. Z., Reininger, K. R., Lancaster, K. M., & Populin, L. C. (2010). Rhesus monkeys (*Macaca mulatta*) do recognize themselves in the mirror: Implications for the evolution of self-recognition. *PLoS ONE*, *5*, e12865. <http://dx.doi.org/10.1371/journal.pone.0012865>
- R Core Team. (2015). *A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria.
- Reiss, D., & Marino, L. (2001). Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 5937–5942. <http://dx.doi.org/10.1073/pnas.101086398>
- Röell, A. (1978). Social behaviour of the Jackdaw, *Corvus monedula*, in relation to its niche. *Behaviour*, *64*, 1–122. <http://dx.doi.org/10.1163/156853978X00459>
- Roma, P. G., Silberberg, A., Huntsberry, M. E., Christensen, C. J., Ruggiero, A. M., & Suomi, S. J. (2007). Mark tests for mirror self-recognition in capuchin monkeys (*Cebus apella*) trained to touch marks.

- American Journal of Primatology*, 69, 989–1000. <http://dx.doi.org/10.1002/ajp.20404>
- Shaffer, V. A., & Renner, M. J. (2000). Black-and-white colobus monkeys (*Colobus guereza*) do not show mirror self-recognition. *International Journal of Comparative Psychology*, 13, 154–160.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9, 76–80. <http://dx.doi.org/10.1177/1745691613514755>
- Soler, M., Pérez-Contreras, T., & Peralta-Sánchez, J. M. (2014). Mirror-mark tests performed on jackdaws reveal potential methodological problems in the use of stickers in avian mark-test studies. *PLoS ONE*, 9, e86193. <http://dx.doi.org/10.1371/journal.pone.0086193>
- Suddendorf, T., & Collier-Baker, E. (2009). The evolution of primate visual self-recognition: Evidence of absence in lesser apes. *Proceedings Biological Sciences*, 276, 1671–1677. <http://dx.doi.org/10.1098/rspb.2008.1754>
- Taylor, A. H., Elliffe, D., Hunt, G. R., & Gray, R. D. (2010). Complex cognition and behavioural innovation in New Caledonian crows. *Proceedings Biological Sciences*, 277, 2637–2643. <http://dx.doi.org/10.1098/rspb.2010.0285>
- Taylor, A. H., Hunt, G. R., Medina, F. S., & Gray, R. D. (2009). Do new caledonian crows solve physical problems through causal reasoning? *Proceedings Biological Sciences*, 276, 247–254. <http://dx.doi.org/10.1098/rspb.2008.1107>
- Uchino, E., & Watanabe, S. (2014). Self-recognition in pigeons revisited. *Journal of the Experimental Analysis of Behavior*, 102, 327–334. <http://dx.doi.org/10.1002/jeab.112>
- Vanhooland, L., Bugnyar, T., & Massen, J. M. (2019). Crows check contingency in a mirror yet fail the mirror-mark test. *Journal of Comparative Psychology*. <http://dx.doi.org/10.1037/com0000195>
- Watanabe, S. (2002). Preference for mirror images and video image in Java sparrows (*Padda oryzivora*). *Behavioural Processes*, 60, 35–39. [http://dx.doi.org/10.1016/S0376-6357\(02\)00094-3](http://dx.doi.org/10.1016/S0376-6357(02)00094-3)

Received June 13, 2019

Revision received February 19, 2020

Accepted March 18, 2020 ■