



LUND UNIVERSITY

Social context hinders humans but not ravens in a short-term memory task

Bobrowicz, Katarzyna; Osvath, Mathias

Published in:
Ethology

DOI:
[10.1111/eth.12971](https://doi.org/10.1111/eth.12971)

2020

Document Version:
Peer reviewed version (aka post-print)

[Link to publication](#)

Citation for published version (APA):
Bobrowicz, K., & Osvath, M. (2020). Social context hinders humans but not ravens in a short-term memory task. *Ethology*, 126(2), 125-139. <https://doi.org/10.1111/eth.12971>

Total number of authors:
2

General rights

Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Social context hinders humans but not ravens in a short-term memory task

Running title: Social context motivates ravens...

Katarzyna Bobrowicz¹, Mathias Osvath¹

¹ Department of Philosophy and Cognitive Science, Lund University, Sweden, Box 192,
22100 Lund

+46728414549, katarzyna.bobrowicz@lucs.lu.se

ORCID 0000-0002-3469-6339 (KB)

ORCID 0000-0002-7873-0930 (MO)

Acknowledgments: We thank Joost van der Weijer (Lund University) for assistance with time-unit kappa and Shortened Complex Span Tasks, and we gratefully acknowledge Lund University Humanities Lab. This work was funded by the Swedish National Council, Grant No. 2014-6402 conjoined with Marie Skłodowska Curie Actions, Cofund, INCA 600398.

Abstract

Using resources shared within a social group – either in a cooperative or a competitive way - requires keeping track of own and others' actions, which, in turn, requires well-developed short-term memory. Although short-term memory has been tested in social mammal species, little is known about this capacity in highly social birds, such as ravens. We compared ravens (*Corvus corax*) with humans in spatial tasks based on caching, which required short-term memory of one's own and of others' actions. Human short-term memory has been most extensively tested of all social mammal species, hence providing an informative benchmark for the ravens. A recent study on another corvid species (*Corvus corone*) suggests their capacity to be similar to the humans', but short-term memory skills have, to date, not been

compared in a social setting. We used spatial set-ups based on caches of foods or objects, divided into individual and social conditions with two different spatial arrangements of caches (in a row or a 3x3 matrix). In each trial, a set of three up to nine caches was presented to an individual that was thereafter allowed to retrieve all items. Humans performed better on average across trials, but their performance dropped, when they had to keep track of partner's actions. This differed in ravens, as keeping track of such actions did not impair their performance. However, both humans and ravens demonstrated more memory-related mistakes in the social than in the individual conditions. Therefore, whereas both the ravens' and the humans' memory suffered in the social conditions, the ravens seemed to deal better with the demands of these conditions. The social conditions had a competitive element, and one might speculate that ravens' memory strategies are more attuned to such situations, in particular in caching contexts, than is the case for humans.

Keywords: short-term memory, raven, human, sociality, primacy, cache recovery

1. Introduction

Well-developed memory systems allow for swift adaptation to complex environments. Keeping track of food locations, actions of conspecifics, and presence of predators and prey has been thought to drive the evolution of diverse memory skills (Murray et al., 2016). Some environments are more cognitively demanding than others and require enhanced memory. For instance, living in a dynamic social group requires well-developed memory because its members must constantly keep track of actions of others to adapt and benefit from group activities, from social interactions (Seyfarth & Cheney, 2015) to foraging (Bugnyar, 2013). This is particularly demanding in groups with fluctuations in group size and group composition (a high degree of fission-fusion dynamics; Aureli et al., 2008, Loretto et al., 2017, Szipl et al.,

2018), such as those of humans (Aureli et al., 2008), some non-human primates (Seyfarth & Cheney, 2015) and some corvids (Bugnyar, 2013; Boucherie et al., 2019).

Limited and ephemeral food availability also taxes memory processing: one must remember when and where food will likely reappear in the future after prolonged periods of absence. Some animals use memory to extend food availability beyond periods of food abundance by hoarding (Clayton & Dickinson, 1998; Vander Wall, 1990; Zinkivskay et al. 2009). Hoarders cache food in several locations during abundance and retrieve it later, when availability drops (Pravosudov & Roth, 2013). The animals need well-developed spatial memories for the location of caches: they must recall where the food was hidden.

Retrieving food from one's own caches is a memory-intensive task, but arguably it is even harder to remember caches in a social context, in which the animal has access to a pool of its own and others' caches, and has to keep track not only of its own actions, but also those of conspecifics. Some social-food hoarders, such as the common raven (*Corvus corax*), resolve this by observing and remembering where a conspecific cached its food to pilfer the cache later when potential competitors are no longer attending (Bugnyar, 2013; Scheid & Bugnyar, 2008). Therefore, to maximize foraging success, ravens need not only individual spatial memory (for their own actions), but also observational spatial memory (for the actions of others; Scheid & Bugnyar, 2008).

Therefore, ravens have to deal with at least two memory-intensive socioecological pressures: high variability in food availability (Vander Wall, 1990) and competing with conspecifics for food caches (Bugnyar, 2013). Under these pressures, ravens have likely acquired enhanced memory skills for food locations and actions of conspecifics (Healy et al., 2005; Pravosudov & Roth, 2013; Smulders et al., 2010), but the limits of these skills remain unclear. To our knowledge, although ravens' long-term memory has been previously

investigated in cognitive tasks (Müller et al., 2017; Boeckle & Bugnyar, 2012), the limits of their short-term memory have not.

This study investigated how ravens' memories are affected by a competitive social context. To gain further knowledge about possible adaptations, we compared the ravens with humans, as humans are not only a highly social species, but also have the best studied memory systems. We used spatial memory tasks, adapted to each species body size and motor system, divided into individual and social settings.

In ecological contexts, ravens typically recover caches within a few hours up to a few days after the caches have been made (Bugnyar, 2013). To store and subsequently retrieve the locations from long-term memory, ravens must first process them within working memory (Baddeley & Hitch, 1974) and/or short-term memory (Atkinson & Shiffrin, 1968). These two terms have been often used interchangeably in animal memory research to refer to cognitive processing of immediate contexts, but they are not identical. Whereas working memory allows for active processing of information from the environment and long-term memory, short-term memory serves as a buffer between these two sources of information (Roberts & Santi, 2017). Processes of working memory prepare information held in short-term memory for later storage in long-term memory (Cowan, 2017), and both working and short-term memory correlate with long-term memory in humans (Neath et al, 2018). Here, we tested the subjects' memory in immediate contexts, and because of (1) the difficulty of establishing to what extent working memory and/or short-term memory contributed to performance in our cache recovery tasks, and of (2) possible terminological confusions, we adopt the term of *short-term spatial memory* used in previous research on cache recovery in corvids (Scheid & Bugnyar, 2008). However, when citing work of others which have used *working memory*, we maintain this term.

Thanks to working memory, an individual can temporally maintain goal-relevant information (Baddeley, 2003; Conway and Engle, 1995) and select it over competing yet less

relevant information. Having a larger storage capacity, and so being able to simultaneously manipulate a larger number of items, has been associated with better performance in general cognitive tasks in humans (Conway and Engle, 1995; Engle et al., 1999; Kyllonen and Christal, 1990; Süß et al., 2002) and mice (Kolata et al., 2005; Light et al., 2010; Wass et al. 2013). In other words, variation in working memory has been related to variation in general intelligence (Kolata et al., 2005), and as such could have a role in cognitive evolution (Beaman, 2010). A well-developed working memory capacity in humans has been argued to have allowed for the emergence and sustenance of sophisticated cognitive abilities—such as planning, innovation, and analogical reasoning (Coolidge & Wynn, 2004).

Humans have repeatedly been reported to have an exceptional working and short-term memory capacity, in general larger than other animals (e.g., Wright & Elmore, 2016, Glassman et al., 1994; Carruthers, 2013). However, healthy adult humans can hold 4 chunks (groups) of items in their short-term memory (Cowan, 2001), and it has recently been shown that crows and rhesus monkeys can hold 4 items in their working memory too (Balakhonov & Rose, 2017; Buschman et al., 2011). Nevertheless, humans are known to effectively use such chunks through different strategies, for instance, by remembering a nine-digit phone number in three three-digit portions. Some animals, such as rats, pigeons and rhesus monkeys, have been shown to use chunking strategies too (e.g., Fountain, 1990; Terrace, 1991; Scarf et al., 2018). However, to our knowledge, this capacity has never been investigated in a social context, in any animal including humans; that is, when one has to update one's own short-term memory based on actions performed by another individual, and act upon the update.

It is difficult to predict how humans and ravens will compare in the spatial memory tasks; if, in such tasks, the raven short-term memory capacity is more limited than the human, the ravens' performance would be worse than that of the humans. Furthermore, humans may use linguistically based mnemonic strategies which are likely unavailable to ravens. As it has

been shown that the ability to solve complex cognitive tasks, that require executive systems, is inhibited by presence of other individuals (Wagstaff et al., 2008), we hypothesize that humans will suffer from a social inhibition effect in our observational spatial memory tasks, but will excel in the individual spatial memory tasks. Further, based on the ravens' socio-ecology, we expect that the ravens might not suffer from a social inhibition effect, and may even perform better in the social than in the individual conditions. Finally, to test whether humans' performance in our setup correlated with their working memory capacity, all humans participated in so-called Span Tasks from Engle Lab (Conway et al., 2005; Foster et al., 2015). This would also indicate what memory functions our tasks measured, which allows for more informed speculations about the performances of the ravens as well.

2. Method

2.1. Subjects

Six adult humans (3 females, $M^{\text{age}}=27.8$, age range: 25-31 years) and six adult ravens (5 females) participated in the study. The humans were tested both individually and in pairs at Lund University, Sweden; they were alumni or current students of the university, and of international background; none of them had a degree in psychology and/or cognitive science. The humans were familiar both with the experimenter and the partner, with whom they participated in two observational spatial memory tasks. The humans were rewarded with cinema tickets, in accordance with the informed consent forms, which were signed before testing. Five out of six ravens were housed at the Lund University Corvid Cognition Station in a social group in a 400 m² space. One raven was a wild free-flying individual, kin to those housed at the Station and voluntarily participating in the experiment. The tests were conducted both individually and in pairs, in familiar facilities with free access to food and water during the experiment. All ravens, including the wild one, were familiar with the experimenter.

The subject-partner pairs were always the same, both in the humans and in the ravens (humans: 1 with 2, 3 with 4, 5 with 6; ravens: Rickard with None, Juno with Embla, Rugga with Tosta). For further details see Supplementary Information 1.

2.2. Apparatus

2.2.1. Ravens

Two experimental set ups were used (Figure 1A-B). Both consisted of nine wooden dishes (8.5x8.5 cm), equidistantly distributed either in a 9x1 row over a long wooden board (113x17x1.2 cm; 2.3 cm distances) or in a 3x3 matrix over a square wooden board (63x63x1.2 cm, 13 cm distances). Although apparently simpler, the row distribution may be less ecologically valid than the matrix distribution as ravens arguably rarely cache in straight rows. Each dish was square-shaped and made by four wood pieces attached to the board. During the experiment, the dishes served as potential cache locations; a food item could be placed within the dish, and then covered with wood chips to disguise the item. The distribution of the dishes ensured that the raven could not explore two caches simultaneously, and that it was forced to walk in front of/between the caches, if more than three were baited in a trial. The food items were always quarters of ring-shaped dog treats.

2.2.2. Humans

Two analogical set ups were used for the humans. However, they were upscaled so that motor effort during cache recovery would be comparable to that of the ravens. This meant that the humans were also forced to walk in front of/between the caches, if more than three were baited in a trial. The humans were confronted with nine wooden dishes (18x18 cm), equidistantly distributed either in a 9x1 row (180x20x20 cm; 7 cm distances) or in a 3x3 matrix (140x140x20cm; 40 cm distances). The dishes were placed over cubes (20x20x20 cm), fastened onto the tops of one-meter high columns that allowed for comfortable cache recovery (without bending; Figure S1). During the experiment, the dishes served as potential cache locations; a

small metal ring (ø1 cm) could have been placed within the boundaries of the dish, and then covered with grill wood chips to disguise the item.

2.3. Procedure

At the beginning of each trial, the experimenter inserted exactly one item (a quarter of the dog treat or a small metal ring) per dish into a number of the dishes, ranging from three to nine. Next, wood chips or stones were placed over the item, and once ensured that it was no longer visible, the experimenter would step back from the apparatus to allow the subject or its partner to approach and immediately start the search. This procedure was followed in all experimental conditions, both with the ravens and the humans.

Each subject participated in four experimental conditions based on four available combinations of two manipulated factors: (1) the set up (row or matrix), (2) the social context (without or with a partner). Depending on the social context, the subjects had to follow different rules when solving the task. Without the partner, the subjects had to follow one rule: “explore each cache one after another, and do not return to an already explored cache”, and if they participated with a partner, they had to follow another rule: “explore caches that have not been explored by the partner”. In the test trials on the row distribution, the experimenter always cached the items in adjacent dishes to maintain a uniform distance between the caches regardless of their number, and to avoid an overlap with training trials. In the test trials on the matrix distribution, the experimenter cached the items in random dishes; otherwise a 3-cache trial on the matrix distribution would be virtually identical to the 3-cache trial on the row distribution. Overall, each subject completed 35 pseudo-randomized trials within each of the four experimental condition (row-individual, matrix-individual, row-social, matrix-social), that is, a total of 140 trials. The subject could not be tested on the same number of caches more than twice in a row, and otherwise the order of cache numbers was pseudo-random. Cache distribution (row vs. matrix) and social context (individual participation vs. participation with

203 a partner) were manipulated. The order of the conditions for each subject was not pseudo-
204 randomized for two reasons. First, the ravens, highly neophobic (Miller et al., 2015), were
205 hesitant to approach the apparatus at the beginning of each condition, regardless of whether
206 they have had access to it beforehand. Because the risk of that a high level of arousal associated
207 with neophobia (Greenberg, 2013) could hinder ravens' performance in the task, a short training
208 phase preceded each condition (see Short-term individual spatial memory tests). Keeping to a
209 predictable order of trials, that is using one set up consistently after the training phase, allowed
210 for reducing the level of arousal and thereby its effect on the ravens' performance. Even when
211 the ravens had already been familiarized with either of the setups in another condition than the
212 immediately preceding one, they reverted to the previous behavior and did not instantly
213 approach the set up. For instance, when presented with the matrix after completing the row-
214 social condition, the ravens still needed a short habituation phase to approach and explore the
215 matrix despite previous experience from the matrix-individual condition. Second, we assumed
216 that the ravens may confuse the rules associated with the individual and the social condition,
217 and we could not be sure that the ravens did not observe each other's trials for two reasons: (1)
218 the testing was not performed in a complete blind spot in the facilities, and could in principle
219 have been observed by the non-participating birds; (2) the wild raven was not housed in the
220 facilities, but was let in during testing, and could possibly observe the testing of other
221 individuals while free-flying or sitting on the top of the aviary. Observing others would not help
222 in increasing memory performance, but as each trial for each individual was unique, the ravens
223 could simply confuse the rules between the individual and the social condition. For this reason,
224 we settled on the same order of conditions for all subjects. We tested the humans in the same
225 manner, and monitored (and quantified) whether the individual subjects improved their
226 performance over the trials and across the conditions. Therefore, all subjects have completed

the tests in the following order: 1. row-individual, 2. matrix-individual, 3. row-social, 4. matrix-social.

2.3.1. Short-term individual spatial memory tests

Both the ravens and the humans participated individually in two short-term individual spatial memory tests, in which they were required to retrieve all hidden items. They were allowed to explore all caches, but not to go back to an already explored cache. When the subject went back to such a cache, the experimenter would immediately interrupt searching and terminate the trial. A cache was coded as explored once the subject touched wood chips within the cache boundaries, even if the subject did not retrieve the hidden item.

In the first individual spatial memory condition, the subjects were confronted with the 1x9 row distribution, and with the 3x3 matrix distribution in the second. Each condition consisted of 35 pseudo-randomized trials, that is, of five trials with each number of caches from three to nine (three or four, or five, or six, or seven, or eight, or nine). In the test trials on the row distribution, the experimenter always cached the items in adjacent dishes, and on the matrix distribution, the experimenter cached the items in random dishes. The row-individual condition was always preceded by three training trials, in which the subject had to retrieve all items from three non-adjacent caches (Figure 2). To proceed to the test, the subject had to reach the criterion of 67% successful trials, which was always reached within the first three training trials. The matrix-individual condition always commenced after the row-individual condition, and followed the same rule, and therefore did not require any training trials. It also solved another issue: because in the matrix-individual condition the caches were always randomly distributed, a training trial with three caches would always be in principle identical to a test with three caches. We expected that both the humans and the ravens would be able to transfer the rules of the task from the row to the matrix distribution because it has previously been shown that corvids are able to transfer abstract rules across tasks (Veit & Nieder, 2013).

In the ravens, the trials were administered in one session per day, in five to ten trials per session. Between the trials, the raven was allowed to leave the experimental setup and cache the retrieved items, and the humans returned the collected items to the experimenter. If the raven did not return to the apparatus within five minutes, the session was terminated for the day. In the humans, the individual trials were always administered on the same day, in a single session per condition, but after ten trials the experimenter asked whether the subject needed a break. There was a ten-minute pause between the row-individual and the matrix-individual condition.

2.3.2. Short-term observational spatial memory tests

After the individual spatial memory tests, both the ravens and the humans participated in pairs in two short-term observational spatial memory tests. They were always paired with the same partner. First, they observed the actions of the partner, and were thereafter allowed to explore the caches left unexplored by the partner. If the subject explored a cache that had been touched by the partner, but still contained a food item, it was allowed to continue the search, either until it retrieved all items or touched an empty cache. Again, the subject was not allowed to touch an already explored cache.

In the first observational spatial memory condition, the subjects were first confronted with the 1x9 row distribution, followed by the 3x3 matrix distribution. Again, each condition consisted of 35 pseudo-randomized trials. At the beginning of the trial the experimenter prepared a certain number of caches, and then stepped back to allow the partner to approach and retrieve a part of the caches: half of them for the even numbers, and a half minus one for the odd numbers. Once the partner finished recovering the last cache, the experimenter would step forward and say “Thank you” to signal that the partner should move away. The raven was let into to another compartment, and the human stepped back to the side of the apparatus. The experimenter ensured that the subject’s view of the apparatus was not obstructed. Next, the

277 experimenter would let the subject approach the apparatus, either by letting it in the
278 experimental compartment (the ravens) or by saying “Please start”/ “You can start now” (the
279 humans).

280 Both social conditions, regardless of the distribution, were preceded by three training
281 trials (Figure 3). In the row-social condition, the subject trained on a total of four non-adjacent
282 caches, and in the matrix-social condition, on a total of two randomly located caches. At the
283 beginning of a trial, the subject could only observe the caches from behind a mesh. The partner
284 gained the access to the caches first, and could explore roughly a half of the caches (1 out of 2
285 or 2 out of 4 in the training trials; 1, out of 3, 2 out of 4, 2 out of 5, 3 out of 6 etc. in the test
286 trials), being observed by the subject. Once the partner explored the allowed number of caches,
287 the subject would be allowed to explore the rest. Again, to proceed to the test, the subject had
288 to reach the criterion of 67% successful trials, which was reached by all subjects within the first
289 three training trials. The matrix-social condition always commenced after the row-social
290 condition.

291 In the ravens, the trials were administered in one session per day, in three to seven trials per
292 session. The subject (observer) and the partner (demonstrator) would switch their roles within
293 a pair unexpectedly across the session to reduce attention lapses. For instance, if the observer
294 completed seven trials in a session, it would first complete three trials, after which it would act
295 as the demonstrator for the other subject for some trials. At some point in the session the roles
296 are again swapped, and so on. Such role-swaps served two purposes: on the one hand, it
297 alleviated the cognitive load because acting as the observer required more attentional resources
298 than acting as the demonstrator; and, the role-swaps made each session different and
299 unpredictable as the swap could occur after any number of trials. Between the trials, the ravens
300 were allowed to leave the apparatus and cache the retrieved items, and the humans returned the
301 collected items to the experimenter. If the observer did not look at the demonstrator’s actions,

the session was terminated for the day. The ravens always completed a single session per day, and, therefore, the overall data collection took about 50 days. In the humans, the social trials were always administered on the same day, in a single session per condition, but after ten trials the experimenter would ask if the subject needed a break. There was a ten-minute pause between the row-individual and the matrix-individual condition. Four subjects completed the individual and the social trials on the same day, with a 2-hour lunch break in between, and two subjects completed the experiment on two separate days.

2.3.3. Shortened Complex Span Tasks

In the humans, the short-term spatial memory tests were preceded by three complex span tasks: operation span (OSpan), symmetry span (SymSpan) and rotation span (RotSpan; Foster et al., 2014). These were computerized tasks, used as a validated measure of human working memory capacity to examine if scores in the spatial memory tasks corresponded to scores in the span tasks. In each of the span tasks, subjects are presented with a sequence of two to seven to-be-remembered items (such as a sequence of letters). Between the presentations of the sequences, subjects have to complete distractor tasks. For instance, in the OSpan tasks, the subject needs to remember sequences of letters, but has to complete simple math problems between the sequences. In addition, after the completion of all span tasks and all spatial memory tasks, the subjects were asked which strategies (if any) they used when solving the spatial tasks.

2.4. Coding

All trials were video-recorded, and for each trial several variables were coded:

1. *Success rate* (a continuous variable), defined as a ratio of correctly touched caches to all caches available in a trial.
2. *Score* (a binary variable), defined as success if all caches were touched without making a mistake.

3. *A- and B-mistakes* (a nominal variable), defined for the individual and the social conditions separately (only in the failed trials).

a. Individual conditions: an A-mistake was coded if the subject returned to a cache that it previously explored, but not immediately preceding the last correct cache, and a B-mistake was coded if it returned to a previously explored cache immediately preceding the last correct cache.

b. Social conditions: an A-mistake was coded if the subject touched a cache previously explored by the partner in its second or later choice, and a B-mistake was coded if the subject touched such a cache in its first choice.

In both conditions, B-mistakes were so simple that they likely resulted from attention lapses rather than memory failures; A-mistakes could result from both shortcomings. If B-mistakes were more frequent than A-mistakes in a given condition and a given species, it suggested that the subjects did not pay attention even to their first choice; on the other hand, if A-mistakes were more frequent than B-mistakes, it would suggest that the subjects in general paid attention to their first choice.

4. *Retention interval* (a continuous variable [in seconds]), defined as an interval between the experimenter's last touch on the last cache and the first touch of the subject's hand/beak on the first cache.

5. *Duration per cache (DPC)*; a continuous variable [in seconds]), defined as an interval between the first touch of the subject's hand/beak on the first cache and the first touch on the last correctly chosen cache, divided by the number of all correctly chosen caches.

6. *First cache explored by the subject* (a nominal variable), coded for the first cache touched by the subject at the beginning of a search.

7. *First cache made by the experimenter* (a nominal variable), coded for the first cache made by the experimenter at the beginning of a trial.

8. *Last cache made by the experimenter* (a nominal variable), coded for the last cache made by the experimenter before the subject approached and started its search.
9. *Overall delay [s]* (a continuous variable), defined as a sum of *Retention Interval [s]* and *Duration per cache [s]* multiplied by a number of correctly touched caches in a given trial.

For the humans, individual overall scores (defined as a number of successes) on each condition were compared with the absolute scores on the span tasks. Although partial scores were also available and are favored over the absolute ones in some situations (Conway et al., 2015), the absolute scores were a better match for the overall scores in the spatial tasks. In the individual and the observational spatial memory tasks, a score of 1 was given only if the subject recovered all available caches, that is, if it was 100% accurate. The criterion of 100% accuracy is also prerequisite for receiving a non-zero absolute score in the span tasks. In our case, to match the partial score, the success rate could be used, but variability in the success rate was much lower than in the score between the subjects in the individual and the observational spatial memory tasks.

The strategies reported by the subjects are available in Table S1.

2.5. Statistics

2.5.1. Fail probability

This experimental setup allowed for different strategies of cache recovery. For instance, the subject could explore the caches randomly, or by using a fixed pattern (e.g., from left to right), or by using memory of the already explored caches; and each of these strategies would lead to different patterns in (1) fail probability over trials, and (2) success rate over trials. For a list of the possible strategies and the corresponding patterns see Table S2.

Whenever the subject chose the caches in a random manner, (1) fail probability should follow a specific pattern, different in the individual and the social condition (Figure 4). These

patterns were compared with those observed, generated by each subject within each condition. In each case two-sample Kolmogorov-Smirnov test (two-sided) was used to determine whether the two patterns were significantly different (ks.boot function from Matching package in R).

2.5.2. Success rate

For all conditions together and for each species separately, general linear mixed-model analysis was used to determine the effects of the number of caches, cache distribution, social context and retention interval on the success rate, controlling for a random effect of subject ID. Beta distribution best fitted the success rate in both the ravens and the humans (glmmTMB function from glmmTMB package in R, Anova function from car package in R). To determine the highest number of caches, after which the ravens' performance significantly dropped, a post-hoc test was performed. Effect sizes were estimated with r2 function from sjstats package in R.

2.5.3. Score

For all conditions together and for each species separately, general linear mixed-model analyses were used to determine the effects of the number of caches, cache distribution, social context and retention interval on the score, controlling for a random effect of subject ID. Binomial distribution best fitted the success rate in both the ravens and the humans (glmer function from lme4 package in R, Anova function from car package in R). Effect sizes were estimated with r2 function from sjstats package in R.

2.5.4. Mistakes

Binomial distribution was a best fit both in the ravens and in the humans (glmer function from lme4 package in R, Anova function from car package in R). Two-sided exact binomial test was subsequently used to determine whether there was a significant difference between a number of A-mistakes and a number of B-mistakes (binom.test in R). Effect sizes were estimated with r2 function from sjstats package in R.

2.5.5. Duration per cache [DPC]

Each subject was allowed to take unlimited time to explore each cache. First, to compare the intervals spent on each cache in each trial between the humans and the ravens, and, second, to determine the effect of the number of caches and the success rate a general linear mixed-model analysis was used, with subject ID as a random variable. Log-normal distribution best fitted the DPC distribution in both species (glmmPQL function from MASS package in R). Effect sizes were estimated with r2beta function from r2glmm package in R.

For each species separately, an additional general linear mixed-model analysis was used to determine the effect of the number of caches, the success rate, cache distribution and the social context on DPC, with subject ID as a random variable (glmmPQL function from MASS package in R). Effect sizes were estimated with r2beta function from r2glmm package in R.

Furthermore, only for the successful trials (score=1) and for each species separately, a general linear mixed-model analysis was used to determine the effect of the number of caches, cache distribution and social context on the DPC (glmmPQL function from MASS package in R). Effect sizes were estimated with r2beta function from r2glmm package in R.

2.5.6. Span tasks vs. spatial memory tasks

To determine whether there was any correlation between the absolute scores in the span tasks and the scores in the spatial memory tasks, linear regression was used (lm function in R).

2.5.7. Serial position effect in individual spatial memory tasks

For each of the individual conditions (row and matrix) and for each species separately, a general linear mixed-model analysis was used to determine the effects of the first and the last cache made by the experimenter on the first cache explored by the subject, controlling for a random effect of subject ID. Gamma distribution best fitted the success rate in both the ravens and the humans (glmmPQL function from MASS package in R, Anova function from car package in R). Effect sizes were estimated with r2beta function from r2glmm package in R.

2.5.8. Overall delay and retention intervals

For each of the individual conditions and for each species separately, a general linear mixed-model analysis was used to determine the effect of the overall delay on the subject's success rate, controlling for a random effect of the subject ID. Beta distribution best fitted the success rate in both the ravens and the humans (glmmTMB function from glmmTMB package in R, Anova function from car package in R).

To compare retention intervals between the two species, a Wilcoxon signed rank test with continuity correction was used (wilcox.exact function from exactRankTests package in R), as the intervals were not distributed normally. Because distributions of this variable were right-skewed in both species, medians (Mdn) and median absolute deviations (MAD) are reported.

2.5.9. Learning effect

For each condition and for each species separately, to test for the learning effect between the second and the first half of trials, a paired Wilcoxon signed rank test with continuity correction was used (wilcox.exact function from exactRankTests package in R).

3. Results

3.1. Fail probability

Whereas none of the humans explored the caches randomly in neither of the experimental conditions, some of the ravens did so, especially in the individual conditions (Figures S2-S3). Further details are provided in Supplementary Information 1.

3.2. Success rate

In the ravens, there was a significant main effect of the number of caches ($\chi^2(6)=39.38$, $p<0.001$, $R^2=0.12$), a significant main effect of cache distribution ($\chi^2(1)=7.12$, $p=0.008$, $R^2=0.74$), and a significant interaction effect of cache distribution and social context ($\chi^2(1)=6.51$, $p=0.011$,

$R^2=0.49$) on the success rate. Specifically, the success rate was significantly higher in 3-cache trials than in 6- ($z=3.099$, $p=0.031$), 7- ($z=4.382$, $p<0.001$), 8- ($z=4.052$, $p<0.001$) and 9-cache trials ($z=4.447$; $p<0.001$) and in 4-cache trials than in 7- ($z=3.797$, $p=0.002$), 8- ($z=3.466$, $p=0.01$) and 9-cache trials ($z=3.866$, $p=0.002$). There were no significant differences in the success rates between trials with 5 or more caches. Furthermore, the success rate was significantly higher in the row individual condition than in the matrix individual (weakly: $z=2.568$, $p=0.05$), row social ($z=3.615$, $p=0.002$) and matrix social ($z=2.673$, $p=0.039$). In the humans, none of these effects were significant (Figure 5).

3.3. Score

In the ravens, there was only a significant main effect of the number of caches on the score ($\chi^2(6)=91.81$, $p<0.001$, $R^2=0.205$; Figure 6A). Specifically, the score decreased as the number of caches increased; the difference in the score was significant between three caches and five to nine caches, between four caches and five to nine caches, and between five and nine caches. In the humans, only a main effect of the social context on the score was significant ($\chi^2(1)=13.27$, $p<0.001$, $R^2=0.208$; Figure 6D). Specifically, the score was significantly higher in the individual than in the social conditions ($z=3.64$, $p<0.001$).

3.4. Mistakes

In the ravens, only a main effect of social context on the mistake type was significant ($\chi^2(1)=11.62$, $p<0.001$, $R^2=0.095$; Figure S4A). Specifically, in the individual conditions, there was no significant difference between a number of A-mistakes (likely memory failures) and B-mistakes (likely attention failures; $p=0.624$), but in the social conditions, there were significantly fewer B-mistakes than A-mistakes ($p<0.001$). In the humans, the main effect of social context was on the verge of significance ($\chi^2(1)=3.51$, $p=0.06$, $R^2=0.233$), likely due to a

limited dataset, as the humans made fewer mistakes than the ravens. However, the same relationships were found for the individual and the social conditions as in the ravens (Figure S4B). Specifically, in the individual conditions, there was no significant difference between a number of A-mistakes and B-mistakes ($p=1$), but in the social conditions, there were significantly fewer B-mistakes than A-mistakes ($p<0.001$).

Both in the ravens and in the humans, the absolute number of B-mistakes was similar regardless of social context. However, in both groups, the absolute number of A-mistakes (likely memory failures) was much higher in the social than in the individual condition (Figures S4C-S4D).

3.5. Duration per cache

For both species together, there was a main effect of species ($\chi^2(1)=7.78$, $p=0.005$, $R^2=0.052$) on the mean DPC (duration per cache), with subject ID as a random variable. There was also an interaction effect of species and the success rate ($\chi^2(1)=6.08$, $p=0.014$, $R^2=0.091$), as well as a weak significant interaction effects of the number of caches and the success rate ($\chi^2(6)=12.32$, $p=0.055$, $R^2=0.098$), and species and the number of caches ($\chi^2(6)=11.07$, $p=0.086$, $R^2=0.094$) on the mean DPC, with subject ID as a random variable.

In the ravens, there was only a main effect of the success rate on the DPC ($\chi^2(1)=21.71$, $p<0.001$, $R^2=0.045$; Figure S5A). Specifically, the mean DPC increased as the success rate increased. In the humans, on the other hand, there was only a main effect of the number of caches ($\chi^2(6)=115.43$, $p<0.001$, $R^2=0.136$; Figure S5B). Specifically, the mean DPC increased as the number of caches increased.

Only for successful trials (score=1) in the ravens, there was a main effect of social context ($\chi^2(1)=7.71$, $p=0.006$, $R^2=0.14$), and an interaction effect of cache distribution and social context ($\chi^2(3)=7.16$, $p=0.007$, $R^2=0.15$). Specifically, the ravens spent significantly more time per cache in the social than in the individual conditions ($z=2.78$, $p=0.006$; Figure S6B),

and in the matrix-social condition than in all other conditions (row-individual: $z=4.11$, $p<0.001$; matrix-individual: $z=2.78$, $p=0.027$; row-social: $z=-5.01$, $p<0.001$; Figure S6A). In the humans, there was also a main effect of social context on the DPC ($\chi^2(1)=6.08$, $p=0.014$, $R^2=0.16$), and an interaction effect of cache distribution and social context ($\chi^2(3)=16.33$, $p<0.001$; $R^2=0.23$). Specifically, the humans spent significantly more time per cache in the individual than in the social conditions ($z=-2.47$, $p=0.014$; Figure S6D), and significantly less time per cache in the matrix-social condition than in all other conditions (row-individual: $z=-3.82$, $p=0.001$; matrix-individual: $z=-3.1$, $p=0.01$; row-social: $z=0.06$, $p=0.05$; Figure S6C).

3.6. Scores on the Span tasks

Interestingly, there was a significant negative correlation between the total score on the span tasks and the total score on the spatial memory tasks ($F(1,4)=18.732$, $p=0.012$; adjusted $R^2=0.653$; Figure S7A). Specifically, the significant negative correlation was found between the total score on the span tasks and the total score on the observational spatial memory tasks ($F(1,4)=10.401$, $p=0.032$; adjusted $R^2=0.78$; Figure S7B), but not the individual spatial memory tasks ($F(1,4)=0.442$, $p=0.543$; adjusted $R^2=-0.126$; Figure S7C).

3.7. Serial position effect

In the ravens that did not recover the caches in a random manner (for details see Supplementary Information 1), there was only a main effect of the first cache made by the experimenter on the first cache explored by the subject, both in the row-individual ($\chi^2(1)=29.296$, $p<0.001$, $R^2=0.197$) and in the matrix-individual condition ($\chi^2(1)=11.232$, $p<0.001$, $R^2=0.11$). Interestingly, this effect was found in the successful (row-individual: $\chi^2(1)=15.751$, $p<0.001$, $R^2=0.182$; matrix-individual: $\chi^2(1)=9.391$, $p=0.002$, $R^2=0.176$), but not in the failed trials (row-individual: $\chi^2(1)=1.789$, $p=0.181$; matrix-individual: $\chi^2(1)=2.463$, $p=0.117$). No effects of the

first and the last made cache on the first cache explored by the subject were found in the ravens that recovered the caches in a random manner.

The humans displayed a similar pattern in their performance. In the humans, again there was only a main effect of the first cache made by the experimenter on the first cache explored by the subject, both in the row-individual ($\chi^2(1)=44.866$, $p<0.001$, $R^2=0.316$) and in the matrix-individual condition ($\chi^2(1)=10.357$, $p=0.001$, $R^2=0.052$). Because there were no failed trials in the row-individual and only two failed trials in the matrix-individual condition, a separate analysis for the failed trials would not be statistically meaningful. The humans that always recovered the caches in a fixed pattern (from left to right) in the row-individual condition were excluded from this analysis.

3.8. Overall delay and retention intervals

In the ravens, there was no effect of the overall delay on the success rate in any of the conditions (row-individual: $\chi^2(1)=2.27$, $p=0.132$; matrix-individual: $\chi^2(1)=3.411$, $p=0.07$; row-social: $\chi^2(1)=1.904$, $p=0.168$; matrix-social: $\chi^2(1)=1.13$, $p=0.288$). Likewise, in the humans, there was no effect of the overall delay on the success rate in any of the conditions (row-individual: $\chi^2(1)=0.028$, $p=0.867$; matrix-individual: $\chi^2(1)=0.015$, $p=0.902$; row-social: $\chi^2(1)=0.06$, $p=0.806$; matrix-social: $\chi^2(1)=0.003$, $p=0.953$; Figure S8).

Interestingly, the median of retention intervals in the ravens was three times longer than in the humans ($W=296140$, $p<0.001$; ravens: $Mdn=12.985$ s, $MAD=11.727$ s, $Max=217.3$ s; humans: $Mdn=4.015$ s, $MAD=1.794$ s, $Max=65.22$ s).

3.9. Learning effect

In both species, there was no significant difference in the scores between the 1st and the 2nd half of the trials in any of the four conditions. For further details see Supplementary Information.

4. Discussion

Only few ravens (four on the row, three on the matrix) used memory outside of the social context, contrary to the humans that always seemed to do so. The ravens' performance did not drop in the social conditions compared to the individual ones, contrary to the humans who demonstrated such a drop. The ravens seemed to use simpler and likely less cognitively demanding strategies in the individual conditions, such as making random choices, whereas the humans (except for two subjects in the row-individual condition) relied on more complex strategies, such as chunking. In the social condition, however, the ravens clearly used more complex strategies. In general, in the social conditions both the ravens and the humans made more memory mistakes (A-mistakes) than in the individual conditions, suggesting that their memory was impaired in presence of others. Neither the ravens nor the humans seemed to reach the limit of their short-term memory, but this requires further studies. Humans, but not ravens, needed more time per cache as the number of caches increased, but neither humans' nor ravens' performance depended on retention intervals. Further, both humans and ravens exhibited the primacy effect in the individual spatial memory tasks, but in the ravens, this was true only for the successful trials. In line with our predictions, scores on the Span tasks correlated with the spatial memory tasks; however, contrary to these predictions, the correlation was negative. We have not detected learning effects within conditions. It is unlikely that such effects occurred across conditions in the ravens, as the success rate was the highest in the first administered condition, that is, the row-individual condition. While this might suggest that the drop in the success rate in the subsequent conditions was caused by e.g., a drop in attention or motivation, it is also unlikely. If such a drop occurred, the ravens should have, for instance, spent less time on each cache in the subsequent conditions. However, this was not the case.

4.1. Primacy effect and retention intervals

Overall, it seems that similar memory processes underlay the ravens' and humans' performance because the primacy effect (better recall for the first caches made by the experimenter) was detected for both species in the individual conditions, and only in trials in which they used memory-based strategies and managed to recover all caches. The primacy effect occurs when the neural network responsible for encoding becomes fatigued, the more items it has to encode, especially if the items are very similar (Tulving, 2008). This effect emerges after relatively long delays (retention intervals) between the presentation of the last item in a series and the onset of the test (Tulving, 2008). Right after the presentation of all items, a subject typically shows the best recall for the last presented items (so-called recency effect). After intermediate delays there are similar levels of recall for the last and the first presented items (so-called intermediate effects). Only after a longer delay the recall becomes better for the first presented items, that is the primacy effect emerges with time. What a relatively long delay is to the subjects differs between species, and is shorter in some animals than in humans (e.g., 10 s in pigeons and 100 s in humans; Wright, 1985). This suggests that the neural network in some animals become fatigued faster, and therefore are prone to loss of more information over time than is the case for humans. However, we found no such differences between ravens and humans in this study. Although the retention intervals were on average longer in the ravens than in the humans, we detected a similar primacy effect in the ravens and the humans. Moreover, we did not find a drop in performance in either of the two species when the overall delay got longer (the interval between the end of experimenter's caching and the end of the subject's search). These results show that the ravens were not more susceptible to information loss over time than the humans, and that the time delays in the tasks had no effect on the differences in absolute scores between the two species.

As delay length cannot explain the differences in overall absolute scores between ravens and humans, there must be other reasons for this disparity; at least four non-mutually exclusive

explanations can be identified: (1) ravens have a limited short-term memory capacity regarding number of items that can be processed as compared to humans; (2) the ravens might differ in attentional capacities; (3) the ravens' motivation was lower than the humans' (resulting in lower attention); (4) the two species used different memory strategies.

4.2. Score, success rate and mistakes

A lower capacity regarding the number of items that can be encoded does, however, not seem to explain the results. A recent study showed that another corvid species (*Corvus corone*) had a working memory capacity of four items (Balakhonov & Rose, 2017), which is similar to the capacity repeatedly shown in humans (Cowan, 2001). The crows and the humans were tested in different setups; however, this similarity gets further support from the current study, as both species were able to keep a stable (flat) success rate when the task became more and more demanding, on the trials with five or more caches. For instance, the humans would perform with 100% accuracy and the ravens with a 65% accuracy on all numbers of caches – that is the ravens *could* remember up to nine caches. A similar sharp drop in performance, but above four items, was also observed in the previously mentioned study on crows. The authors explained this as a result from changes in motivation, or a possible difference in short-term memory mechanisms between corvids and primates (Balakhonov & Rose, 2017).

It is likely that there was a drop in attention in the ravens when the number of caches exceeded five. Interestingly, the ravens exhibited more memory failures in the social conditions than in the individual conditions, but the number of attention failures was similar regardless of the context. This suggests that the presence of others was detrimental for both humans' and ravens' memory performance, even though such presence did not seem to affect the ravens' scores in the task. This suggests that, contrary to the humans, the ravens could deal with the demands of the social conditions without suffering a drop in overall gain from the task.

There are also some indications of differences in attention/motivation and strategies between the ravens and the humans. A comparison between the groups, based solely on the absolute scores, is not straightforward. In fact, different factors were responsible for the drop in the absolute score in the two species: the ravens' scores got lower as the number of caches increased, and the humans' scores got lower when they participated with the partner.

4.3. Duration per cache

The humans and the ravens dealt differently with the demands of increasing numbers of caches. Humans spent gradually more time (on average) on the individual caches. Ravens, on the other hand, spent the same average amount of time per cache within a trial, regardless of the number of caches involved. However, the average time spent on caches differed between trials, and in the trials in which they were more successful, the average time per cache was higher than in the unsuccessful ones. In other words, the humans kept high accuracy levels over trials, but got slower as the task got more difficult. The ravens, however had lower accuracy over trials, but did not get slower on average within trials when difficulty increased.

The human tradeoff between speed and accuracy could be a result of linguistic processing, which allows for complex strategies such as assigning abstract symbols (numbers) to the caches, which at the same time slows down the performance. Indeed, the two humans that had the highest scores in all conditions, reported to have used linguistic strategies. Interestingly, the same subjects reached the lowest overall scores on the computerized working memory (Span) tasks, which cannot be encoded with such strategies. This might suggest that these subjects had learned to compensate their limitations in working memory with pronounced linguistic strategies. This shows the importance of such strategies for short-term spatial memory, and how it can buffer limitations of the working memory; however, this relationship calls for further investigation in the future.

The differences within the ravens between successes and fails, seems to be best explained by attention and motivation. As mentioned, even if there was no difference of average time spent per cache within trials, the average time differed between trials and correlated with success: the longer, the better. That is, the ravens could be as successful as the humans if they spent more time per cache, but they did not always do this. Perhaps because it required more expended effort than in the humans, and that this effort was not motivated by the gains, which might be true even if the effort was not greater than in humans (the ravens got rewards in every trial anyway, and rarely consumed all of them). That the difference between time spent per cache per trial was a result of motivational factors becomes clearest in the social conditions.

The ravens spent more time on the caches in the social condition than in the individual conditions, indicating a higher motivation spurred by the social context. Interestingly, the humans did the opposite, and spent less time in the social conditions. The most pronounced differences between the species were found in the matrix condition, where the ravens spent more time and the humans spent less time than in any other condition.

4.4. Strategies

That the motivation increased for the ravens in the social conditions seems evident, but it also suggests that they used different strategies from the individual conditions, or used more complex strategies more often in the social conditions. Ravens compete for resources with conspecifics and spend a lot of time caching and recaching food when potential competitors are present (Bugnyar, 2013). The matrix social condition appears to be the most ecologically valid condition from this perspective, which might explain the use of strategies requiring more time per cache.

The humans, on the other hand, might have been hindered in their strategies by someone else being involved in the task, which is indicated by decrease both in time spent per cache as

well as in overall success. When asked for their strategies, they reported that they used one strategy for all conditions as long as possible, such as operating on numbers assigned to individual caches or planning the order of search before they approached the setup. This was obviously less effective in the social conditions. Theoretically, the subjects could also have used another strategy: of remembering, which caches are empty and not which are still baited. Although none of the humans reported this strategy, it might have been used by the ravens.

This study cannot clearly disentangle what strategies were used more precisely by the ravens in the social conditions, which makes any comparisons with the human strategies difficult. It is however reasonable to assume that ravens might have predisposed memory strategies for a competitive caching context. If one were to further speculate, it could be the case that human short-term memory in social contexts is more attuned to cooperative task, while the opposite might be true for ravens. This requires further studies.

Neither the ravens nor the humans seemed to reach the limit of their short-term memory in our tasks because they could keep track of up to nine items in all conditions. To complete the tasks, the subjects had to represent the number of the caches, maintain it on-line in working memory and execute accurate movements following cognitive processing (e.g., inhibit going back to the already explored caches). Nine caches go beyond the working memory capacity in both species, so to succeed with remembering them some memory strategies must have been used. Using such strategies, in turn, requires representing the number of caches, and updating the representation during the search. As it has previously been shown that corvids can represent numericities ranging from 1 to 30 (Ditz & Nieder, 2015; Ditz & Nieder, 2016), the ravens should have been able to represent the varying numbers of caches and therefore could have used memory strategies, which may partly explain their performance.

Further studies should compare short-term memory performance between the species in cooperative tasks. Our setup could be adapted to such tasks. For instance, subjects could

recover tokens instead of food rewards, and only if the subject and its partner collected a full set of tokens, the set could be exchanged for food rewards divided equally between the subject and the partner. This would not only allow for a comparison between short-term memory performance in a cooperative setup between the species, but could increase the overall levels of motivation in all conditions, increase the level of attention in the social conditions and limit provisions for fake recoveries in the social conditions. Moreover, corvids and other food-hoarding animals could be further tested in setups that do not rely on cache recovery; such studies would reveal whether the effect of social context is confined only to predisposed domains.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. The facility and the care taking routines were approved by the Swedish Agricultural Board (No 5.2.18-5395/16). Ethical approval for the procedures was granted by the regional ethics board for animal research in the county of Skåne (No M 333-12).

5. References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation*, 2, 89-195. doi:10.1016/S0079-7421(08)60422-3

727 Aureli, F., Schaffner, C. M., Boesch, C., Bearder, S. K., Call, J., Chapman, C. A., . . . Schaik,
 728 C. P. v. (2008). Fission-Fusion Dynamics: New Research Frameworks. *Current*
 729 *Anthropology*, 49(4), 627-654. doi:10.1086/586708
 730 Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and*
 731 *Motivation*, 8, 47-89. doi:10.1016/S0079-7421(08)60452-1
 732 Baddeley, A. (2003). Working memory looking back and looking forward. *Nature Reviews*
 733 *Neuroscience*, 4, 829–839. doi: 10.1038/nrn1201
 734 Balakhonov, D., Rose, J. (2017). Crows Rival Monkeys in Cognitive Capacity. *Scientific*
 735 *Reports*, 7, 8809. doi: 10.1038/s41598-017-09400-0
 736 Beaman, C.P. (2010). Working memory and working attention: What could possibly evolve?.
 737 *Current Anthropology*, 51, S27–S38. doi: 10.1086/650297
 738 Boeckle, M., & Bugnyar, T. (2012). Long-Term Memory for Affiliates in Ravens. *Current*
 739 *Biology*, 22(9), 801-806. doi:10.1016/j.cub.2012.03.023
 740 Boucherie, P. H., Loretto, M.-C., Massen, J. J. M., Bugnyar, T. (2019). What constitutes
 741 “social complexity” and “social intelligence” in birds? Lessons from ravens.
 742 *Behavioral Ecology and Sociobiology*, 73, 12. doi: 10.1007/s00265-018-2607-2
 743 Bugnyar, T. (2013). Social cognition in ravens. *Comparative Cognition & Behavior Reviews*,
 744 8, 1-12. doi:10.3819/ccbr.2013.80001
 745 Buschman, T. J., Siegel, M., Roy, J. E., & Miller, E. K. (2011). Neural substrates of cognitive
 746 capacity limitations. *Proceedings of National Academy of Sciences of the United*
 747 *States of America*, 108, 11252-11255. doi: 10.1073/pnas.1104666108
 748 Carruthers, P. (2013). Evolution of working memory. *Proceedings of the National Academy*
 749 *of Sciences of the United States of America*, 110(2), 10371-10378.
 750 doi:10.1073/pnas.1301195110

751 Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by
752 scrub jays. *Nature*, 397(17), 269-274.

753 Conway, A. R., & Engle, R. W. (1995). Working memory and retrieval: A resource-
754 dependent inhibition model. *Journal of Experimental Psychology: General*, 123, 354–
755 373.

756 Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R.
757 W. (2005). Working memory span tasks: A methodological review and user's guide.
758 *Psychonomic Bulletin & Review*, 12(5), 769-786.

759 Coolidge, F. L., & Wynn, T. (2004). A cognitive and neuropsychological perspective on the
760 Châtelpéronian. *Journal of Anthropological Research*, 60, 55-73.

761 Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental
762 storage capacity. *Behavioral and Brain Sciences*, 24, 87–114. doi:
763 10.1017/S0140525X01003922

764 Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic*
765 *Bulletin & Review*, 24, 1158-1170. doi: 10.3758/s13423-016-1191-6

766 Ditz, H. M., Nieder, A. (2015). Neurons selective to the number of visual items in the corvid
767 songbird endbrain. *Proceedings of the National Academy of Sciences of the United*
768 *States of America*, 112(25), 7827-7832. doi: 10.1073/pnas.1504245112

769 Ditz, H. M., Nieder, A. (2016). Sensory and Working Memory Representations of Small and
770 Large Numerosities in the Crow Endbrain. *The Journal of Neuroscience*, 36(47),
771 12044-12052. doi: 10.1523/NEUROSCI.1521-16.2016

772 Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory,
773 short-term memory, and general fluid intelligence: A latent variable approach. *Journal*
774 *of Experimental Psychology: General*, 128, 309–331.

775 Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W.
 776 (2015). Shortened complex span tasks can reliably measure working memory capacity.
 777 *Memory & Cognition*, 43(2), 226-236. doi:10.3758/s13421-014-0461-7
 778 Fountain, S. B. (1990). Rule abstraction, item memory, and chunking in rat serial-pattern
 779 tracking. *Journal of Experimental Psychology: Animal Behavior Processes*, 16, 96-
 780 105. doi: 10.1037/0097-7403.16.1.96
 781 Glassman, R. B., Garvey, K. J., Elkins, K. M., Kasal, K. L., & Couillard, N. L. (1994). Spatial
 782 Working Memory Score of Humans in a Large Radial Maze, Similar to Published
 783 Score of Rats, Implies Capacity Close to the Magical Number 7+-2. *Brain Research*
 784 *Bulletin*, 34(2), 151-159.
 785 Greenberg, R. (2003). The Role of Neophobia and Neophilia in the Development of
 786 Innovative Behaviour of Birds. In S. M. Reader & K. N. Laland (Eds.), *Animal*
 787 *Innovation*: OUP Oxford.
 788 Güntürkün, O., Bugnyar T. (2016). Cognition without Cortex. *Trends in Cognitive Sciences*,
 789 20(4), 291-303. doi: 10.1016/j.tics.2016.02.001
 790 Healy, S. D., Kort, S. R. d., & Clayton, N. S. (2005). The hippocampus, spatial memory and
 791 food hoarding: a puzzle revisited. *Trends in Cognitive Sciences*, 20(1), 17-22.
 792 doi:10.1016/j.tree.2004.10.006
 793 Kolata, S., Light, K., Townsend, D. A., Gregory Hale, Grossman, H. C., & Matzel, L. D.
 794 (2005). Variations in working memory capacity predict individual differences in
 795 general learning abilities among genetically diverse mice. *Neurobiology of Learning*
 796 *and Memory*, 84, 241-246. doi:10.1016/j.nlm.2005.07.006
 797 Kyllonen, P. C., & Christal, R. E. (1990). Individual differences in associative learning and
 798 forgetting. *Intelligence*, 12, 389-421.

799 Light, K. R., Kolata, s., Wass, C., Denman-Brice, A., Zagalsky, R., & Matzel, L. D. (2010).
800 Working memory training promotes general cognitive abilities in genetically
801 heterogeneous mice. *Current Biology*, 20, 777–782. doi:10.1016/j.cub.2010.02.034

802 Loretto, M. C., Schuster R., Itty, C., Marchand, P., Genero, F., & Bugnyar, T. (2017). Fission-
803 fusion dynamics over large distances in raven non-breeders. *Scientific Reports*, 7, 380.
804 doi: 10.1038/s41598-017-00404-4

805 Miller, R., Bugnyar, T., Pölzl, K., & Schwab, C. (2015). Differences in exploration behaviour
806 in common ravens and carrion crows during development and across social context.
807 *Behavioral Ecology and Sociobiology*, 69, 1209-1220. doi:10.1007/s00265-015-1935-
808 8

809 Müller, J. J. A., Massen, J. J. M., Bugnyar, T., & Osvath, M. (2017). Ravens remember the
810 nature of a single reciprocal interaction sequence over 2 days and even after a month.
811 *Animal Behaviour*, 128, 69-78. doi:10.1016/j.anbehav.2017.04.004

812 Murray, E., Wise, S., & Graham, K. (2016). *The Evolution of Memory Systems*: OUP Oxford.

813 Neath, I., Saint-Aubin, J., Bireta, T. J., Gabel, A. J., Hudson, C. G., & Surprenant, A. M.
814 (2019). Short- and Long-Term Memory Tasks Predict Working Memory Performance,
815 and Vice Versa. *Canadian Journal of Experimental Psychology*, 73(2), 79-93. doi:
816 10.1037/cep0000157

817 Pravosudov, V. V., & Il, T. C. R. (2013). Cognitive Ecology of Food Hoarding: The
818 Evolution of Spatial Memory and the Hippocampus. *Annual Review of Ecology,*
819 *Evolution, and Systematics*, 44, 173-193. doi:10.1146/annurev-ecolsys-110512-
820 135904

821 Roberts, W. A., & Santi, A. (2017). The comparative study of working memory. In J. Call
822 (Ed.), *APA Handbook of Comparative Psychology* (Vol. 2. Perception, Learning, and
823 Cognition): American Psychological Association.

824 Scarf, D., Smith, C. D., Jaswal, V. K., Magnuson, J. S., & Terrace, H. (2018). Chunky
825 Monkey? The spontaneous temporal chunking of simultaneous chains by Humans
826 (Homo Sapiens) and rhesus monkeys (Macaca mulatta). In *Studies of Rhesus Monkeys*
827 *and their Behaviors*: Nova.

828 Scheid, C., & Bugnyar, T. (2008). Short-term observational spatial memory in Jackdaws
829 (Corvus monedula) and Ravens (Corvus corax). *Animal Cognition*, 11(4), 691-698.
830 doi:10.1007/s10071-008-0160-5

831 Seyfarth, R. M., & Cheney, D. L. (2015). Social Cognition. *Animal Behaviour*, 103, 191-202.
832 doi:10.1016/j.anbehav.2015.01.030

833 Smulders, T. V., Gould, K. L., & Leaver, L. A. (2010). Using ecology to guide the study of
834 cognitive and neural mechanisms of different aspects of spatial memory in food-
835 hoarding animals. *Philosophical Transactions of the Royal Society B*, 365, 883-900.
836 doi:10.1098/rstb.2009.0211

837 Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working
838 memory capacity explains reasoning ability - and a bit more. *Intelligence*, 30, 261-
839 288.

840 Szípl, G., Ringler, E., & Bugnyar, T. (2018). Attacked ravens flexibly adjust signalling
841 behaviour according to audience composition. *Proceedings of The Royal Society B*,
842 285(20180375), 1-9.

843 Terrace, H. (1991). Chunking During Serial Learning by a Pigeon: I. Basic Evidence. *Journal*
844 *of Experimental Psychology*, 17(1), 81-93. doi: 10.1037//0097-7403.17.1.81

845 Tulving, E. (2008). On the law of primacy. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn
846 (Eds.), *Memory and mind: A festschrift for Gordon H. Bower* (pp. 31-48). Mahwah,
847 NJ, US: Lawrence Erlbaum Associates Publishers.

848 Vander Wall, S. B. (1990). *Food Hoarding in Animals*: University of Chicago Press.

- Veit, L., & Nieder, A. (2013). Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. *Nature Communications*, 4, 2878. doi:10.1038/ncomms3878
- Wagstaff, G. F., Wheatcroft, J., Cole, J. C., Brunas-Wagstaff, J., Blackmore, V., & Pilkington, A. (2008). Some cognitive and neuropsychological aspects of social inhibition and facilitation. *European Journal of Cognitive Psychology*, 20(4), 828-846. doi:10.1080/09541440701469749
- Wass, C., Pizzo, A., Sauce, B., Kawasumi, Y., Sturzoiu, T., Ree, F., . . . Matzel, L. D. (2013). Dopamine D1 sensitivity in the prefrontal cortex predicts general cognitive abilities and is modulated by working memory training. *Learning and Memory*, 20, 617-627. doi:10.1101/lm.031971.113
- Wright, A. A., & Elmore, L. C. (2016). Pigeon visual short-term memory directly compared to primates. *Behavioural Processes*, 123, 84-89. doi:10.1016/j.beproc.2015.09.002
- Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., Cook, R. G. (1985). Memory Processing of Serial Lists by Pigeons, Monkeys, and People. *Science*, 229, 287-289.
- Zinkivskay, A., Nazir, F., & Smulders, T. V. (2009). What-where-when memory in magpies (*Pica pica*). *Animal Cognition*, 12(1), 119-125. doi:10.1007/s10071-008-0176-x

Figure captions

Figure 1. A display of experimental set ups for the ravens. Two cache distributions were used: a. a 9x1 row, b. a 3x3 matrix.

Figure 2. A display of training (A-C) and test (D-F) trials in the row-individual condition. In the training, always three single items were inserted into three nonadjacent caches (A), and then covered with wood chips and/or stones (B). The subject was then expected to retrieve all three items (C). In the test, several single items, here five, were inserted into several adjacent

caches (D). Once they were completely covered with wood chips and/or stones (E), the subject was expected to retrieve all hidden items (F).

Figure 3. A display of training (A-C) and test (D-F) trials in the row-social condition. In the training, always four single items were inserted into four nonadjacent caches (A), and then covered with wood chips and/or stones (B). The partner was then allowed to retrieve two items, and only then the subject was allowed to retrieve the remaining items (C). In the test, several single items, here five, were inserted into several adjacent caches (D). Once they were completely covered with wood chips and/or stones (E) and the partner retrieved a half of the items, the subject was expected to retrieve the remaining, in this case three, items (F).

Figure 4. A display of fail probabilities for three-item trials in the (A) individual and the (B) social conditions. (A) In the individual condition, upon the first choice all caches contain an item, and so the first choice is always correct. The second choice is also always correct because the subject can either keep exploring the same cache or choose another out of the two that still contain an item. Upon the third choice, the subject has two options: an empty, already explored cache or a cache that has not been yet explored. Therefore, fail probability in the third choice equals 50%. (B) In the social condition, upon the first choice, one cache has already been emptied by a partner, which means that a chance of choosing this cache is 33%. Upon the second choice, only one non-empty cache is left, which means that there is 50% chance of an incorrect choice.

Figure 5. A display of a main effect of task on the success rate in the ravens (A) and in the humans (C), and an interaction effect of the cache distribution and the social context in the ravens (B) and in the humans (D).

Figure 6. A display of significant main effects on the score in the ravens (A), and in the humans (D). There was a main effect of the number of caches on the score in the ravens, and

898 a main effect of the context in the humans. However, there was no main effect of the number
899 of caches in the humans (B), and no main effect of the context in the ravens (C).