University of Veterinary Medicine Vienna

Messerli Institute

Institute of Animal Welfare Science

Master's Thesis

Investigating Batch Effects in the Use of a Cooperative Task in Pigs (*Sus scrofa*)

Submitted by:

Sophia Maria Weijler, BSc

Matr. Nr.: 01636121

UI 066 222 Interdisciplinary Master in Human-Animal Interactions

Vienna, July 2023

Supervisors: Dr. Jim McGetrick & Univ.-Prof. Dr. Jean-Loup Rault

Plagiarism Disclaimer

I hereby declare that this thesis is my own and autonomous work. All sources and aids used have been indicated as such. All texts either quoted directly or paraphrased have been indicated by in-text citations. Full bibliographic details are given in the reference list. This work has not been submitted to any other examination authority.

Vienna, 28.07.2023

place/date signature

Acknowledgements

First and foremost, I want to express my deepest gratitude to my supervisors, Dr. Jim McGetrick and Prof. Dr. Jean-Loup Rault, for their guidance and constant support throughout the journey of completing this master's thesis. Your expertise and constructive feedback were instrumental in shaping this research. I am extremely grateful for your patience and encouragement throughout the whole process.

I want to thank the head of the department, Prof. Dr. Ludwig Huber, for giving me the opportunity to study and conduct research in such an interesting and academically nurturing environment.

I am grateful for the opportunity to work at the Medau Research Farm and want to express my gratitude towards Dr.med.vet. Doris Verhovsek and the rest of the team at the farm, for their help and support throughout the data collection.

I am very thankful for the help I have gotten from Christian Haberl, with any technological problems and handiwork, as well as for the help from Janina Weißenborn with the reliability coding.

I also want to thank Tessa Parker and Kimberly Brosche, for their help during data collection and their valuable insight and emotional support. I had a great time while working with you!

Finally, I want to thank my friends and my family, who endured my endless monologues about this thesis and supported me in many different ways. Thank you for always being there for me!

Zusammenfassung

Kooperation, welche als gemeinsames Handeln zweier Individuen definiert werden kann, ist bei einer Vielzahl von Spezies untersucht worden, wobei häufig Paradigmen wie der " loose string-pulling task " verwendet wurden. Studien über Kooperation bei Nutztieren sind selten, obwohl die Zusammenarbeit als Indikator für das Wohlbefinden oder sogar als Instrument zur Verbesserung des Wohlbefindens vorgeschlagen wurde. Um die Kooperation bei Nutztieren wie Schweinen zu untersuchen, wurde von Rault und Kollegen im Jahr 2020 der "Joint-Log-Lift task" entwickelt. Diese Aufgabe kann von den Schweinen mit Hilfe ihres natürlichen Futtersuchverhaltens gelöst werden, d. h. durch Wühlen und Heben mit der Schnauze, und kann im Stall präsentiert werden, was freie Partnerwahl ermöglicht. Vorhergehende Studien kamen zu widersprüchlichen Ergebnissen hinsichtlich der Frage, ob Schweine dieses neue Paradigma spontan lösen können. Um herauszufinden, ob das Paradigma als stabile und robuste Methode zur Untersuchung der Kooperation bei Nutztieren eingesetzt werden kann, wurde versucht, eine Folgestudie von McGetrick und Kollegen zu replizieren, in der 70% der teilnehmenden Schweine in der Lage waren, den "Joint-Log-Lift task" sowie den "Individual-Log-Lift task", welcher ähnlich zum "Joint Log Lift task", aber alleine lösbar ist, ohne Training zu lösen. In dieser Studie wurde, mit Hilfe derselben Methoden wie in der vorherigen Studie, zwei neuen Gruppen mit jeweils 24 Schweinen die Möglichkeit gegeben, den "Joint-Log-Lift task" und den "Individual-Log-Lift task" zu erlernen. Die Schweine beider Gruppen waren nicht in der Lage, den "Joint-Log-Lift task" spontan zu lösen, und die Replikation war daher nicht erfolgreich. Um die Gründe für das Scheitern der Replikation zu erörtern, wurde das Verhalten der Schweine der zwei Studien um die Box verglichen. Die Gruppen dieser Studie zeigten ein geringeres Engagement und eine geringere Beharrlichkeit bei dem "Joint-Log-Lift task" als die Gruppe der replizierten Studie. Dies könnte ein Hinweis auf Gruppenunterschiede in Bezug auf Bewältigungsstile oder soziale Beziehungen sein, was auf einen messbaren Gruppeneffekt auf das kooperative Verhalten beim "Joint Log Lift task" hinweist. Die Tatsache, dass die vorherige Studie nicht repliziert werden konnte, unterstreicht die Notwendigkeit, die Reproduzierbarkeit von Studienergebnissen zu prüfen, und verdeutlicht die Bedeutung von kumulativem Wissen und die Notwendigkeit weiterer Erkenntnisse zu erlangen über Faktoren, die die Reproduzierbarkeit beeinflussen.

Abstract

Cooperation, which can be defined as two individuals acting together, has been studied in a variety of animal species, often utilizing paradigms like the loose string-pulling task. Studies on cooperation in farm animals are scarce, even though cooperation has been suggested to serve as an indicator of welfare, or even as a tool to enhance welfare. In order to study cooperation in farm animals like pigs, the Joint-Log-Lift task was designed by Rault and colleagues in 2020. This task can be solved by the pigs using their natural foraging behaviour, i.e. rooting and lifting with their snouts, and can be presented in their home pens, allowing for free partner choice. Previous studies produced opposing findings with regards to whether pigs can spontaneously solve this new paradigm. To determine whether the paradigm could be used as a stable and robust method to study cooperation in farm animals, a follow-up study by McGetrick and colleagues was attempted to be replicated, in which 70% of pigs were found to be able to solve the Joint-Log-Lift task, as well as the Individual-Log-Lift task that is similar but can be solved by one individual alone, without training. In this study, two new batches, consisting of 24 pigs each, were given the opportunity to learn to use the Joint-Log-Lift box, as well as the Individual-Log-Lift box, using the same methods as in the previous study. The pigs of both batches were unable to spontaneously solve the JLL task, and the replication was therefore not successful. To elucidate the reasons for the failure to replicate, the behaviour of the pigs around the boxes in the different studies was compared. The batches of this study showed less engagement with and less persistency at the Joint-Log-Lift task than the batch of the replicated study. This could be indicative of batch differences regarding coping styles or social relationships, indicating a measurable batch-effect on cooperative behaviour in the Joint-Log-Lift task. The failure to replicate the previous study emphasises the need to test the reproducibility of findings and accentuates the importance of cumulative knowledge and the need for further insight into factors influencing replicability.

Table of Content

Introduction

Cooperation, which has been defined as "the simultaneous or consecutive acting of two or more individuals by the same or different behaviours" (Taborsky and Taborsky 2015), has been studied extensively throughout different fields of biology and in various contexts (Dugatkin 2002; Massen et al. 2019; West et al. 2007). The vast body of literature on cooperation shows that it is a phenomenon readily found in many different social species, ranging from social arthropods (Yip et al. 2008; Rodrigues et al. 2022), birds (Hatchwell 2009; Nolen and Lucas 2009) and fish (Lönnstedt et al. 2014; Wong and Balshine 2011) to mammals (Smith et al. 2012; Clutton-Brock et al. 2004). Cooperation occurs in various contexts (Dugatkin 1997). For example, ants have been shown to cooperatively transport food and other resources (Czaczkes and Ratnieks 2013), as well as participate in inter-group cooperation (Rodrigues et al. 2022). Different species of birds have been observed practicing cooperative predator defence using mobbing behaviour (e.g.: noisy miners (Kennedy et al. 2009); chaffinch (Marler 1956); carolina chickadees, tufted titmice, white-breasted nuthatches (Nolen and Lucas 2009)), and 9% of all known bird species have been found to engage in cooperative nest care (Cockburn 2006). Lastly, one of the most prominent examples of cooperation, cooperative hunting and foraging, can be found in wolves (Muro et al. 2011), lions (Stander 1992) and even fish (Lönnstedt et al. 2014).

Cooperation studies often rely on observing non-human animals (henceforth referred to as "animals") in their natural habitat. But, to examine the proximate mechanisms underlying cooperation and to study different aspects and constraints of cooperation, research has to be conducted in a more controlled environment and needs to go beyond observational studies (Noë 2006; Taborsky and Taborsky 2015). For this purpose, various paradigms have been successfully implemented (Massen et al. 2019). One of the most well-known and frequently applied methods used to study cooperation is the loose string-pulling paradigm. This task consists of a reward, oftentimes placed on a platform behind a fence, just out of reach of the studied animals, and a string, with which the studied animals can pull the reward towards them. The string is loosely threaded through loops fixed on the platform with both ends of the string within the reach of the animals. The ends of the string are positioned in such a way, that it is impossible for one animal to pull both at the same time. If only one animal pulls on one end of the string, it comes loose, and the task becomes unsolvable. But if two animals pull at both ends at the same time, the platform is drawn closer to the fence separating the animals from the rewards and comes either into reach of the animals through the fence or slides into the enclosure of the animals through a small opening in the fence. The rewards can then be obtained by the cooperating animals (Jacobs and Osvath 2015). This paradigm has been used to study the capacity for and the underlying mechanisms of cooperation in a variety of animals (e.g.: wolves and domestic dogs (Marshall-Pescini et al. 2017), elephants (Plotnik et al. 2011), rooks (Seed et al. 2008), kea (Schwing et al. 2020), chimpanzees (Hirata and Fuwa 2007) and African grey parrots (Péron et al. 2011)). However, the loose string-pulling task utilises a behaviour that is highly unusual to show for many animals, since pulling is often not a part of their natural behavioural repertoire. In order to study cooperation in a setting that is ecologically relevant and reflects and investigates the natural occurrence of cooperation, paradigms that focus on behaviour naturally shown by the studied animals are preferable (Massen et al. 2019).

Despite the aforementioned vast body of literature on cooperation in various species, studies on cooperation in farm animals are scarce, which limits our understanding of the underlying mechanisms and the importance of it in the lives of farm animals, as well as of the potential usefulness of cooperation in the context of welfare (Rault et al. 2021). Even though studies on the affective and physiological effect of cooperation are seldom conducted and often difficult to interpret (reviewed in Marshall-Pescini et al. 2017; Massen et al. 2019), a study by Dale and colleagues (2020) connected success in a cooperation task to social factors (e.g.: affiliation, strength of social bond), which could substantiate the notion that cooperation could serve as an indicator of welfare and could even be used to actively improve the welfare in farm animals, such as domestic pigs (*Sus scrofa domesticus*) (Rault 2019).

Domestic pigs are highly social animals, building stable relationships with conspecifics (Fels et al. 2012; Gonyou 2001) and sometimes engaging in cooperative behaviour like communal care for their offspring when kept in groups (Fraser et al. 1995). Their ancestors, wild boars (*Sus scrofa*), have been found to live in socially complex groups, which benefit from cooperative foraging (Focardi et al. 2015). In 2020, Rault and colleagues developed a new paradigm to study cooperation in pigs, the Joint-Log-Lift task (JLL task (Rault et al. 2020)). The JLL task uses a box, in which food rewards are placed. Access to those rewards is blocked by a log, which can be moved upwards along drawer slides and will lock in place at the top of the box if lifted high enough. Two pigs need to simultaneously lift this log using their snouts, in order to each reach the food reward located behind the log. The JLL task utilizes pigs' natural foraging behaviour, which mainly involves rooting through the ground with their snouts, while providing the opportunity to offer the task in the pigs' home environment, resulting in minimal disturbance of the group and the possibility of free partner choice (Rault et al. 2021).

In previous studies, pigs have been shown to be able to spontaneously solve the JLL task. In a study by Rault and colleagues (2021) approximately 70% of tested juvenile pigs were able to successfully solve the task within 10 days of having access to the box without previous training. In a follow-up study, McGetrick and colleagues (*in prep.*) found juvenile pigs to be similarly successful in solving the JLL task without training. In their study they aimed to determine whether pigs understood the need for a partner to cooperate, and introduced a second box as control condition, the Individual-Log-Lift box (ILL box), which is almost identical to the JLL box, but can be solved by an individual pig alone. The first batch of pigs participating in this study was exposed to a slightly revised model of the box as used in the study by Rault et al. (2021; rounded logs as opposed to square ones, added railings, pipes as reward inlet and panel, see Fig. 1). This first batch was not able to reliably solve the JLL task successfully. The experimenters theorized that the height the log had to be lifted to could have been too difficult for weaker or smaller pigs to solve. Therefore, they altered the JLL box, adjusting the height the log needed to be lifted to and reran the study with a second batch of pigs. This second batch was then the highly successful batch of pigs with which the rest of the study was carried out. Therefore, even though the study reports juvenile pigs being highly capable of solving the JLL task without previous training, this success was only achieved after adjusting the JLL box. Furthermore, the first published study utilizing the JLL task tested Kune Kune pigs using a similar design of the JLL box as the study by Rault and colleagues (2021). However, these pigs were also not able to solve the task spontaneously when given access to the box. The pigs were only able to reliably solve the task after extensive one-on-one training (Koglmüller et al. 2021).

The performance of the pigs in these three previous studies is difficult to compare and generalize, given that they differ either in the model of the task, as the design of the JLL box was altered for the study of McGetrick et al. (*in prep.*), or housing conditions and training approach. The JLL task could offer new opportunities to expand our insight into cooperation in farm animals and to find new ways to enhance their welfare (Rault 2019). However, it is not clear how robust this method is, given that the studies on the JLL task so far produced conflicting results.

In this study, it was attempted to replicate the findings by McGetrick et al. (*in prep.*) in order to assess the reproducibility of the finding that juvenile pigs confronted with the JLL task are reliably successful in solving this task. In order to do so, the settings of the previous study were mimicked, exposing juvenile pigs to both the JLL box and the ILL box. Furthermore, this study aimed to gain insights into which factors could influence the success of the different batches of pigs. This study consists of two parts, with the first one focusing on the replication of the previous study, and the second part examining the results of the replication study more closely.

Part 1 – Replication Study

1 Material and Methods

This study was conducted from April to May and in June of 2022 over the course of 20 working days each month (excluding weekends) at the Medau Research Farm, Berndorf, Austria, which is owned and operated by the University of Veterinary Medicine, Vienna, Austria. The study was planned to consist of a learning phase, in which the pigs were given the opportunity to spontaneously learn to use the boxes, as well as a testing phase, should the learning phase be successful.

The study which was attempted to be replicated was conducted by McGetrick and colleagues in July 2021 at the same farm.

1.1 Ethical Approval

This study was approved by the Ethics and Animal Welfare Committee of the University of Veterinary Medicine, Vienna, in accordance with the University's guidelines for Good Scientific Practice (ETK-036/02/2022; ETK-094/05/2022).

1.2 Animals

The animals taking part in this study were domestic pigs (*Sus scrofa domesticus,* Large White×Pietrain breeds), which were housed at the Medau Research Farm. Four groups consisting of six pigs each took part in this study, resulting in a total of 24 pigs distributed evenly across four pens. The piglets in each group were taken from six different sows at the age of 5 weeks, ensuring that all piglets in one group were genetically unrelated. Each group consisted of 3 males and 3 females. To be able to distinguish between the pigs, they were marked with livestock marker spray. The pigs were housed under commercial like conditions. Their pens had partially slatted floors and were cleaned daily by the experimenter or the animal keepers. Part of the pigs' enclosure was a floor-heated sleeping area with litter consisting of a straw and sawdust mixture. The animals had *ad libitum* access to drinking water and commercial feed, via a drinking fountain and a multi-spaced automatic feeder. In addition to the sawdust and straw, the pigs were provided with enrichment in the form of to two small orange balls (Dogs Creek Ball Airflow M-L; diameter, 7.6 cm; cat no. 1337241; MULTIFIT

Tiernahrungs GmbH, 47809 Krefeld, Germany), which were cleaned daily, and two jute ropes (thickness: 20 mm) hanging at the back and front of the enclosure.

Our housing conditions were almost identical to those of the previous study which was attempted to be replicated. The only difference was that in this previous study 36 pigs participated instead of 24. However, the group size and composition (3 females and 3 males per group) was the same as in our study.

1.3 Study Overview

This study was planned to consist of two phases: the learning phase and the testing phase, for a physiological study that was not carried out in the end. During the learning phase the pigs were first habituated to the experimenter, the food, and the two boxes, Joint-Log-Lift box (JLL box) and Individual-Log-Lift box (ILL box) and then given the opportunity to learn to use the boxes. During the learning phase the pigs were also habituated to be fitted with and to wear a heart rate belt, for the testing phase, which was ultimately not carried out. Every part of this study was videotaped using videorecorders (HIKVISION, 4MP Darkfighter Box Network Camera, DS-2CD5046G0-AP) overseeing the pens of the study animals.

1.3.1 Joint-Log-Lift Box

The JLL task uses the Joint-Log-Lift box (JLL box; Fig.1). This box consists of a wooden corpus (57.2 cm × 75.4 cm × 43.7 cm) with an acrylic sheet (~4 mm thick) attached to the front, which has two cut outs as openings at the front (width = 15 cm, distance between the openings = 20 cm). Through these openings the reward bowls inside the box and the rewards in the bowls can be accessed. Blocking the openings at the front is a log (length = 67.5 cm; diameter = 10 cm), which can be lifted by the pigs to obtain access to the rewards (lifting height = 14.5 cm, from the floor of the box to the bottom of the log). The log moves along vertical drawer slides on the inside of the box, which protrude though a small cut out at the top of the box, when lifted high enough. At this cut out is a spring-loaded bolt latch, which holds the drawer slide attached to the log in place and prevents the log from falling down again, after it has been lifted high enough. The log can then still be moved higher but will not move lower than the minimum lifting height. Only when the experimenter releases the latch, the log will fall down, and the pigs can try again. The box has railings at the sides (~29 cm long), so the pigs working at the task will not be as easily disturbed by others. A dowel is horizontally inserted inside the wooden corpus, behind the acrylic sheet, to prevent smaller pigs from stepping over the log and gaining access to the reward without solving the task. The box also has two inlets on the top, through which the experimenter can replace the rewards in the bowls. A wooden panel painted in blue can be moved up and down between the acrylic sheet and the wooden corpus of the box. This wooden panel can be used to block access to the log, when the pigs are not supposed to have the opportunity to solve to the task yet.

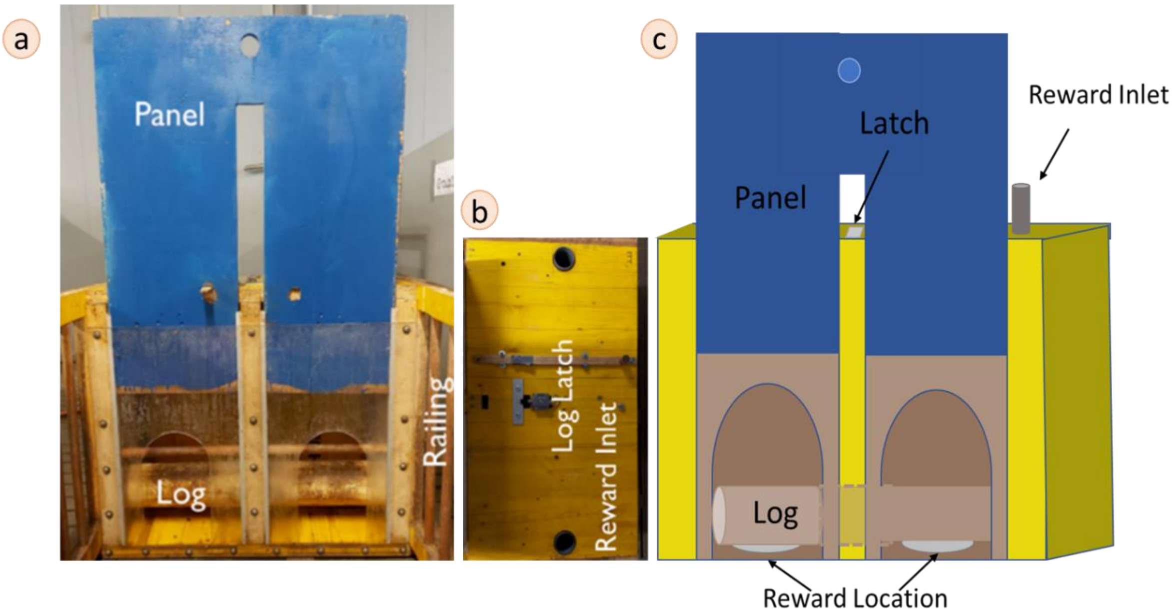

Figure 1: *Joint-Log-Lift box. Actual box used in this study depicted from the front (a), showing the panel, the log and the railing, and from the top (b), showing the reward inlets and the log latch. Schematic outline depicting the reward inlets, the latch, the panel, the reward locations, and the log (c).*

The JLL task can only be solved if two pigs lift the log simultaneously. The log is then fixed in its highest position by the latch, and the pigs can get the reward. If only one pig lifts the log, the log will not properly move along the drawer slides, but rather tilt to one side and the drawer slides will jam in such a way, that the log cannot be moved up further until the pig retreats and the log falls down again (Fig. 2). After a successful lift, and after the rewards are taken, the experimenter can lower the front panel before they reset the log and replace the food reward. They can then open the panels, and the pigs can try again.

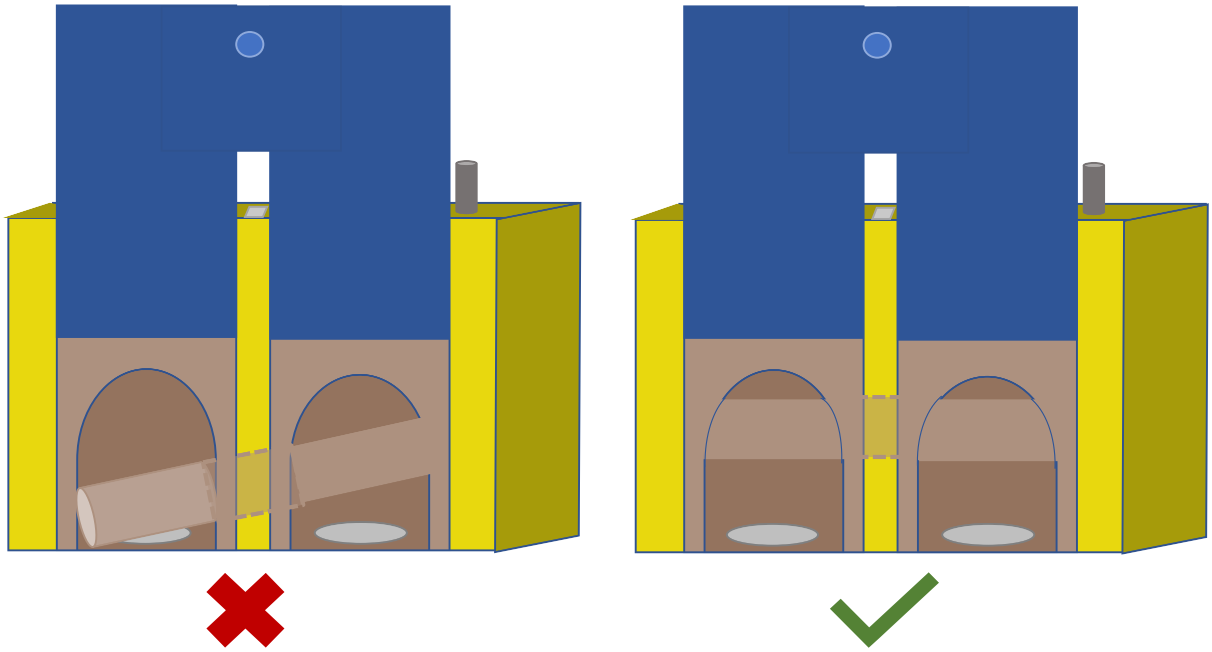

Figure 2: *Schematic picture of a single pig unsuccessfully (left) and two pigs successfully (right) using the Joint-Log-Lift box.*

1.3.2 Individual-Log-Lift Box

The Individual-Log-Lift box (ILL box) is similar to the JLL box (Fig. 3). However, the ILL box has two shorter logs (length = 30.5 cm, diameter = 10 cm), one in front of each reward location, instead of one log blocking both. Furthermore, the box has two wooden panels instead of one panel. Therefore, access to the task can be prevented separately for each log in the ILL box. Given that the reward bowls inside the box are blocked by two separate logs, the pigs can solve this task on their own and do not need to cooperate.

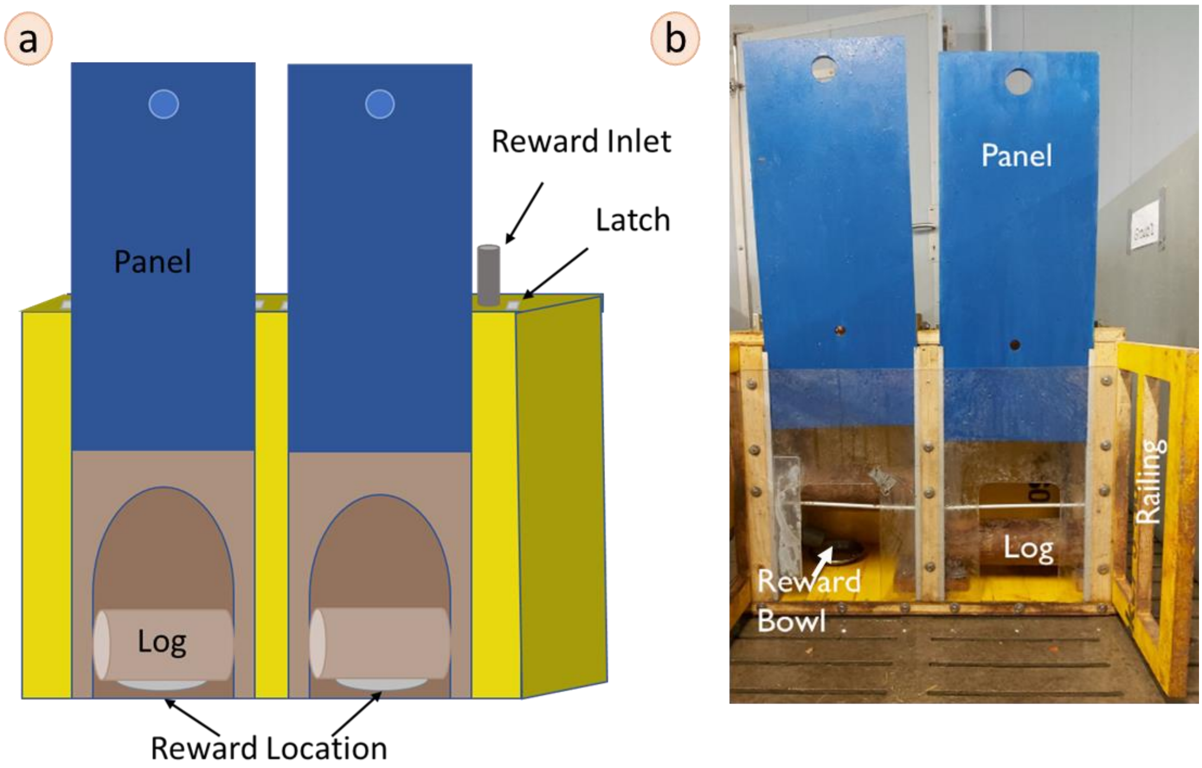

Figure 3: *Individual-Log-Lift box. Schematic depiction of the ILL box (a) and actual box used in this study, depicting reward bowl (log up), log down, railings and panels (b).*

1.3.3 Learning Phase

The learning phase consisted of three stages: habituation to the food, habituation to the boxes, and learning to use both boxes. In total, the learning phase was planned to last 16 days (Fig. 4).

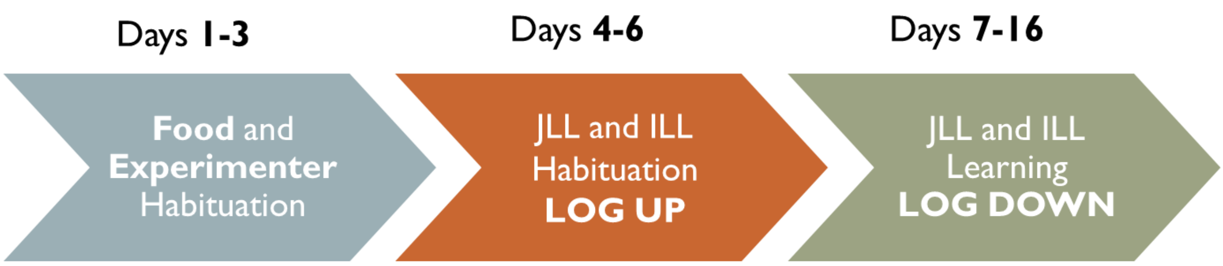

Figure 4: *Timeline of the planned learning phase. On days 1 to 3, habituation to the food and the experimenter took place (see section 1.3.3.1). On days 4 to 6 habituation to the two boxes took place (see section 1.3.3.2). On days 7 to 16 the pigs were given the opportunity to learn to use the boxes (see section 1.3.3.4).*

1.3.3.1 Habituation to the Experimenter and Food

On the first three days of the learning phase, the pigs were habituated to the experimenter and to the food used as a reward, which in this study, were apple pieces (each ~1cm^3). The experimenter spent a total of 15 minutes per day with each group during this phase. The order in which the individual groups were visited by the experimenter was continuously changed by starting with a new group each day. The experimenter entered the enclosure and sat down on the floor. Most pigs still tended to not approach the experimenter upon the first encounter, so the apple pieces were first scattered on the floor to get them accustomed to the taste of apples. When the pigs were bold enough to approach the experimenter, the apple pieces were fed by hand and the experimenter began to stroke the back and the belly of the pigs, to habituate them to physical contact with humans. The experimenter fed approximately one apple cut in small pieces per group and day. How long it took for the pigs to approach the experimenter, and how long it took until they appeared comfortable with physical contact with the experimenter differed between the individual pigs and groups. Therefore, the scattering of the apple pieces was performed for a longer period of time, i.e. for more days, in some groups than in others before the experimenter switched to hand feeding the apple pieces. After the pigs were habituated to the experimenter and the food, the experimenter continued to spend 15 minutes each day with each group of pigs, but they did not receive apple pieces from the experimenter anymore. In this time, habituation to the heart rate belt took place (see section *1.3.3.3 Habituation to the heart rate belt*).

1.3.3.2 Habituation to the Boxes

After the three days of habituation to the experimenter and the food were completed, the pigs were gradually habituated to both the JLL box and the ILL box. The habituation to the boxes took place over the course of three days. In this phase, the boxes were put in the enclosures of the pigs for 15 minutes each, starting with a different box and group every day. The exposure to the boxes consisted of two rounds. In the first round, each group was exposed to one of the boxes. After each group was exposed to one of the boxes, the second round was started. During this second round, each group was given access to the box that was not put in their enclosure in the first round. The order of which group started with which box was changed each day, starting with a different group each day and alternating between which box each group was exposed to in the first round (alternating between groups: e.g. starting with group 1 on day 1, then starting with group 2 on day 2; alternating between boxes: e.g. exposing group 2 to the JLL box first on day 1, then exposing group 2 to the ILL box first on day 2). The box was put into the enclosures with the back against the door at the front, so the experimenter could operate the boxes from outside the enclosure by leaning over the door. The panels in front of the boxes were closed until a timer started to ensure each group had access to the task for the same amount of time. The boxes were placed in the enclosure of the pigs with the logs fixed in their lifted position and the bowls baited before the panels were opened. The pigs had free access to the apple pieces located in the bowls inside the boxes. The experimenter dropped one apple piece per side every 30 seconds through the bait inlets on top of the boxes.

1.3.3.3 Habituation to the Heart Rate Belt

On the days that the pigs were habituated to the boxes, and on the days that they learned to use the boxes, the pigs were also habituated to the heart rate belt. This habituation took place during the 15-minute sessions that the experimenter spent every day with each group after all of the box sessions were finished for the day. The pigs were gradually habituated to the belt using a procedure resembling the habituation procedure recently used by Byrd et al. (2020). The belt was first offered to the pigs to sniff, then dragged across their backs and when they appeared comfortable with that sensation, the belt was fixed around the belly of the pigs, close to their front legs. The time frame of this procedure differed between the individual pigs depending on their timidity and for some individuals it took several days until they appeared to be comfortable with the belt being fitted around their belly. The belt was then left on each pig for about one to two minutes each day. The pigs did not receive any food reward during this habituation.

1.3.3.4 Learning to Use the Boxes

On the following ten days, the pigs were given the opportunity to learn how to use the boxes. The boxes were presented to the pigs in the same manner as during the habituation to the boxes (see section *1.3.3.2*), except that the logs were down, and the experimenter did not add apple pieces every 30 seconds, but the reward bowls were pre-baited with apples pieces. If a log was lifted successfully, and the pigs finished eating the reward (two apple pieces per bowl), the panel in front of the box was closed, the log reset and the reward replaced (two apple pieces per bowl), after which the panel was opened, and the pigs could lift the log again.

The initial performance of the pigs over the first two days of this learning phase was compared with the performance of the study from July 2021(McGetrick et al. *in prep.*). The batch of this current study showed drastically lower success rates than the previously successful batch of 2021, in both the Joint-Log-Lift task (2021: 131 lifts by 27 pigs; 2022: 0 lifts), as well as the Individual-Log-Lift task (2021: 399 lifts by 24 pigs; 2022: 95 lifts by 3 pigs). Therefore, to encourage the pigs of this study to lift the logs in the boxes, the learning phase was altered.

1.3.4 Altered Learning Phase

Given that the success rate was very low, the learning phase was altered after the pigs were exposed to the boxes with the logs down and the bowls baited for two days (Days 7-8; Fig. 5).

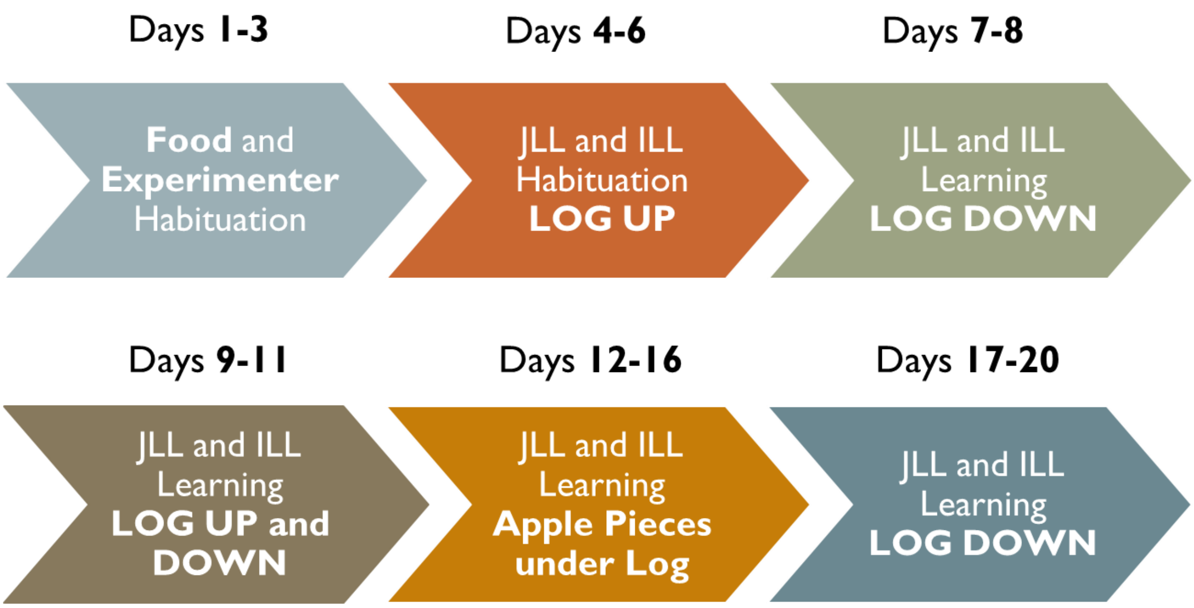

Figure 5: *Timeline of the altered learning phase. The first 6 days were carried out according to the original plan (see Fig. 4). On days 7 to 8 the pigs were given the opportunity to learn to use the boxes with the logs down. On days 9 to 11 the logs were fixed in the upper position for three minutes in the beginning and then down again for 12 minutes. During the days 12 to 16 the logs were down but the experimenter placed apple pieces under the log every 3 minutes. During the days 17 to 20 the logs were down.*

On the next three days (Days 9-11), the pigs were given access to the box for three minutes with the logs locked in their highest position. The experimenter dropped one apple piece per side every 30 seconds, after which the log was lowered again for the next 12 minutes. The aim of this procedure was to remind the pigs that there were apples pieces in the bowls of the boxes and to motivate them to lift the logs in order to obtain more rewards.

The pigs still had no success in lifting the log in the JLL task (0 lifts) and only little success in solving the ILL task (max. 51 lifts per day).

On the following five days (Days 12-16), the boxes were again presented to the pigs with the logs down, and the reward bowls pre-baited. Given that the log was in its lowest position, they did not have free access to food rewards. However, the experimenter put additional apple pieces directly under the log every three minutes. The aim of this procedure was to engage the pigs and prompt them to put their snout under the logs. After eating the rewards under the logs, the pigs were expected to be more motivated to retrieve the pieces from the bowls as well, leading to them lifting the logs. During this period of the habituation phase the majority of pigs learned to solve the ILL task (20 out of 24 pigs lifted the logs more than five times on one day). However, only some pigs managed to lift the log in the JLL box sporadically (a total of 13 lifts in 10 days by 3 pigs in Group1, 5 pigs of Group3 and 2 pigs of Group4). The learning phase was therefore extended by four days (Days 17-20). In these last four days, the boxes were presented with the logs down, and no apple pieces under the logs.

Even though most pigs mastered the ILL task, only some pigs were able to lift the log in the JLL box and no pigs were able to solve the JLL task reliably (see numbers in pervious paragraph).

1.4 Rerun of Replication

In order to rule out sample size as a confounding factor, as only 24 pigs participated in this study in April 2022 as opposed to 36 pigs in the study, which was attempted to be replicated, the decision was made to test a second batch of 24 pigs. This second run was conducted in June 2022. The methods used to rerun the study with this second batch of pigs were identical to the first batch. The pigs were first subjected to the original learning phase. As the success rate was similarly low as in the first batch, the same altered procedure as applied to the previous batch was applied here.

1.5 Behaviour Coding

The video material used for statistical analysis consisted of the first ten days on which the pigs were given the opportunity to learn to use the boxes (Day 7 until and including Day 16, see Fig. 5) from the two batches of this study, as well as the batch from the study of McGetrick et al. (*in prep.*). The individual batches are henceforth called: Batch A (April 2022, first replication study), Batch B (June 2022, second replication study) and Batch C (July 2021, original study being replicated). The video material was analysed using the software BORIS (Friard and Gamba 2016). The behaviour "successful lift" was coded, which was defined as the instance when a pig lifts a log successfully in the ILL task or when two pigs lift the log successfully in the JLL task.

Reliability coding was conducted by comparing the results of two experimenters analysing 30 percent of the first two days of the video material. The videos to be analysed were randomised by assigning numbers to the videos and picking 30 percent of those videos using a number randomizer.

1.6 Statistical Analysis

Statistical analysis was conducted in R statistical software (version 4.2.3 "Shortstop Beagle", R Core Team 2023). Random slopes were identified before running the models and overdispersion was checked where necessary using diagnostics functions kindly provided by Roger Mundry. Analyses resulting in p-values smaller than 0.05 were deemed significant.

1.6.1 Successful Lifts

To compare the success of the pigs in the different batches and to investigate whether performance improved across days, a generalised linear mixed model (GLMM) was fitted with a Poisson distribution using the glmer() function in the "lme4" package (version 1.1.32, Bates et al. 2015). The response variable was the number of successful lifts at the group level, and the fixed effects were Day, Batch and Task, as well as a squared term for the factor Day, to allow for the model to be fitted to a curvilinear relationship between successful lifts and days. The model included a three-way interaction between Day, Batch and Task, as well as between the squared Day factor, Batch and Task (Day*Batch*Task + Day^2*Batch*Task). The factor Day was z-transformed to a mean of 0 and a standard deviation of 1 prior to fitting the model to help with model convergence and interpretation of the results. The factors Group and Observation were included as random intercept effects. One Observation was classified as the gathered data on one pig from one group on one specific day when confronted with one specific box. The random slopes of Day and Task were included within the random effect of Group, with an interaction between Day and Task. The factor Task was manually dummy coded and centred before adding it as a random slope. The model also included a control option, which was specified before fitting the model using the glmerControl() function. The “optimizer” argument was set to “bobyqa”, and the maximum number of function evaluations was set to 100,000.

The model was not over-dispersed (dispersion parameter = 0.8345). However, the estimates in the model output had some extreme values (ranging from -17 to 17). A second model was fitted, in which the squared term for the fixed effect Day, and the correlations among the random slopes and random intercepts were removed. This model was not over-dispersed (dispersion parameter = 0.5819). A full-null model comparison was conducted using the function anova() with the argument “test” set to “Chisq”, with the null model lacking the fixed effects but being otherwise identical to the full model. The model was plotted using base R and, given that the full-null model comparison was significant, levels of the fixed effects were compared by visual inspection of the plot.

Interobserver reliability for the coding of the successful lifts in BORIS was assessed using the Interclass Correlation Coefficient applying the function icc() in the “irr” package (version 0.84.1, Gamer et al. 2019) with the argument “model” set to “twoway” and the argument “type” set to “consistency”. Interobserver reliability was excellent (ICC = 0.99, $p < 0.001$, $n_{Observation} = 18$, $n_{Raters} = 2$).

2 Results

The full-null model comparison revealed that there was a significant effect of the fixed effects, Day, Batch and/or Task, and/or their interactions (χ^2 = 107.8208, df = 11, p < 0.001; Fig. 6). Visual inspection of the plotted model and its 95% confidence intervals indicates that the number of times the pigs of Batch A lifted the log(s) was significantly lower than the number of successful lifts of Batch C in both the ILL task and the JLL task. The number of successful lifts at the ILL task of Batch B seems to be similar to the success of Batch C, while the number of successful lifts of Batch B at the JLL task was significantly lower than the number of successful lifts of Batch C. There seems to be an effect of day for all three batches in the ILL task, while this effect is only strongly prominent in Batch C in the JLL task. However, there seems to be a slight learning curve for Batch B in the JLL task as well, even if a lot less marked. The effect of task is visible for all three batches but is most pronounced in the batches A and B, with strikingly lower numbers of successful lifts at the JLL task than the ILL task.

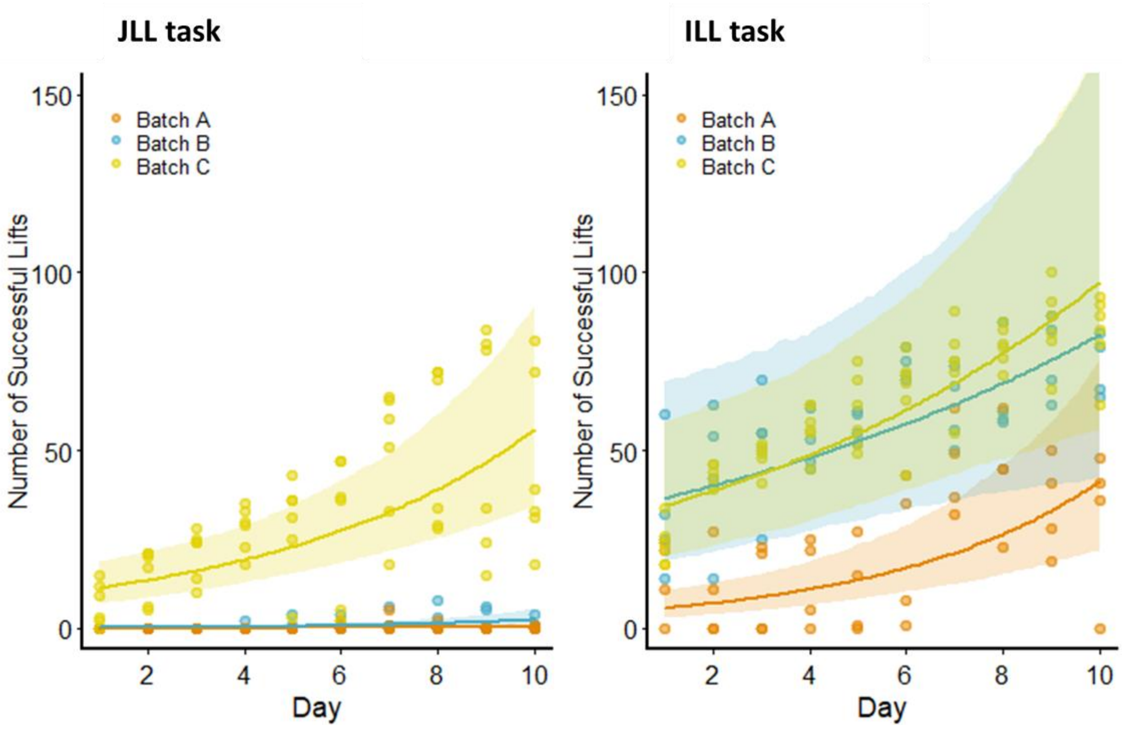

Figure 6: *Number of successful lifts per batch while having access to the JLL task (left) and the ILL task (right). Number of successful lifts depicted for the batches of the current study, Batch A (April 2022, orange), Batch B (June 2022, blue) and the batch of the study of McGetrick and Colleagues (*in prep.*), Batch C (July 2021, yellow) including 95% confidence intervals (colour bands) and individual datapoints.*

3 Discussion

The two batches from this current study were significantly less successful in the JLL task and had only little success in lifting the log in this task (13 lifts over 10 days for Batch A and 50 lifts over 10 days for Batch B vs. 1,956 lifts over 10 days in Batch C). The study which was attempted to be replicated showed a very high success rate, with about 80% of pigs solving the JLL task on their own (McGetrick et al., *in prep*). In this current study the exact same boxes, location, species, and methods were used, and the groups were formed in the same way as in the previous study. But even though the circumstances of the two studies were next to identical, none of the pigs of the current study were able to solve the JLL task reliably and a significantly lower number of pigs from Batch A were able to solve the Individual-Log-Lift task (ILL task). It is therefore unclear why it was not possible in this study to replicate the findings of the study from McGetrick and colleagues (*in prep.*).

Replication is a critical issue, which is slowly starting to gain attention in different scientific fields (Baker, 2016; Farrar et al., 2020). Many studies on replicability focus on methodological and statistical issues and how to refine those aspects. However, studies with seemingly identical methods can still be influenced by factors that are difficult to detect while running the study (Frias-Navarro et al. 2020; Farrar et al. 2020). One of these factors could be the so-called batch-effect. This effect describes behavioural and genetic fluctuations across batches of animals, including fluctuations in individual or interpersonal behaviour, among others (Fuiman et al. 2005; O'Bryant et al. 2011).

The difference in behaviour between the less successful batches A and B of this study and Batch C of the replicated study was further explored to gain insights into the reasons for differing performances across batches. These differences in performances could have been brought upon by differences in interest in the boxes as well as differences in persistence at the given tasks. These differences could exist in the time spent interacting with the boxes, i.e. the engagement the individual batches showed with the boxes, and the attempts to lift the log(s), i.e. the persistence of the different batches at trying to solve the tasks (Massen et al. 2013; Svartberg 2002).

Part 2 – Investigating the Batch-Effect

1 Material and Methods

This follow up study is based on the video material obtained while working with the less successful batches in 2022 (Batch A and Batch B), as well as on the videos from the study conducted in 2021 (Batch C) (McGetrick et al. *in prep.*).

1.1 Animals

The batches A and B consisted of 24 pigs each, while Batch C consisted of 36 pigs. The housing conditions were identical in both studies (see section *2.1 Animals*).

1.2 Behaviour Coding

The focus of this study lies on the level of engagement and persistence shown in the batches from 2022 (Batch A and B) and the batch from 2021 (Batch C) when interacting with the JLL box and the ILL box. The first two days (days seven and eight of the altered learning phase, see Fig. 5; henceforth called Day 1 and Day 2) in which the pigs were given the opportunity to learn to use the boxes with the log(s) down were analysed. Only the first two days were analysed, given that on those days the way in which the boxes were presented to the pigs was identical for both the current study and the study by McGetrick et al. (*in prep.*). In the current study an altered learning phase was implemented after two days in order to try and heighten success rates (see Part 1 section *1.3.4*), given that the success of the batches A and B in solving the two tasks was only a fraction of the success of Batch C. This altered learning phase was not used in the study by McGetrick et al. (*in prep.*). The video material was analysed using the software BORIS (Friard and Gamba 2016). The behaviours coded and analysed are interacting with the front part of or putting the head inside the ILL box or the JLL box, as well as attempting to lift the log(s) and successful log lifts (Table 2).

Reliability coding was conducted by comparing the results of two experimenters analysing 30 percent of the total video material. The videos to be analysed were randomised by assigning numbers to the videos and picking 30 percent of those videos using a number randomizer.

Table 1: *Ethogram of coded and analysed behaviour of pigs when interacting with the ILL box and the JLL box.*

Behaviour	Description
Interacting with front part of box	Interacting with the front of the box using the snout or the mouth, including chewing on the Plexiglas. Does not apply when pig puts head inside cut out at front of box. "Head in box" stops "interacting with front of box" and vice versa.
Head in box	Pig puts head past the ears inside cut out at front of box, runs even when successful lift.
Attempt to lift	Pig interacts with log inside box, puts snout underneath it and tries to lift it (successful + unsuccessful ones are counted)
Successful lift	Pigs lifts the log, which is then locked in the highest position by the latch

1.3 Statistical Analysis

Statistical analysis was conducted in R statistical software (version 4.2.3 "Shortstop Beagle", R Core Team 2023). Random slopes were identified before running the models and overdispersion was checked where necessary using diagnostics functions kindly provided by Roger Mundry. Factors included in the fitted models were dummy coded and centred prior to inclusion as random slopes and the covariate Day was always z transformed prior to inclusion as a fixed effect or random slope.

Interobserver reliability for the coding of all of the aforementioned behaviours was carried out as described in section *1.6 Statistical Analysis.*

Analyses resulting in p-values smaller than 0.05 were deemed significant. All of the data was plotted using the ggplot() + geom_boxplot() function of the gglpot2 package (version 3.4.2, Wickham 2016).

1.3.1 Engagement

1.3.1.1 Interacting with Front of the Box

The effect of Day, Task and Box on the proportion of time pigs spent interacting with the front of the box was analysed. The proportion of time pigs spent interacting with the front of the box was calculated by combining the duration of "interacting with front part of box" and "head in box" (as both values reflect engagement with and possible interest in the boxes and the tasks) and dividing it by the total amount of time the pigs had access to each box (i.e.: 15min.). The resulting variable was called "interacting with front of box". The analysis was carried out by fitting a generalized linear mixed model (GLMM) with a beta distribution and a logit link function using the glmmTMB() function in the "glmmTMB" package (version 1.1.7, Brooks et al. 2017). As the data included a lot of zeros, and the beta regression cannot handle zeros, the response variable was transformed, using a normalization transformation, which readjusts the values to range between zero and 1 (0 < values > 1) in order to fit the model. Day (1, 2), Batch (A, B, C) and Task (JLL, ILL) were included as fixed effects and Subject and Group were included as random intercept effects. A three-way interaction between fixed effects Day, Batch and Task was included. The random slopes of Day and Task were included within the random effects of Subject and Group.

As an overall test of the effect of Day, Batch and Task on the time spent interacting with the front of the box, the full model was compared to a null model (Forstmeier and Schielzeth 2011). The null model did not include any fixed effects but was otherwise identical to the full model. The comparison was conducted using the function anova() with the argument "test" set to "Chisq". The p-values for the fixed effects were obtained using the function drop1 with the "test" argument set to "Chisq". None of the interactions were significant, therefore the model was run again, without the three-way interaction. The model was not over-dispersed (dispersion parameter = 0.772). The full-null model comparison and the extraction of the p-values for this new model without the interactions was conducted in the same manner as described above.

Interobserver reliability was excellent (ICC = 0.891, $p < 0.001$, $n_{Observations} = 18$, $n_{Raters} = 2$)

1.3.2 Persistence

1.3.2.1 Attempt to Lift

To analyse the effect of Day, Batch and Task on the total amount of attempts to lift the log(s), including attempts that led to a successful lift, a zero-inflated GLMM was fitted using the mixed_model() function in the package "GLMMadaptive" (version 0.8-8, Rizopoulos 2023). The response variable was all attempts to lift (successful and unsuccessful), and the fixed effects were Day, Batch and Task. A three-way interaction was also included between the random effects (Day*Batch*Task). The zero-inflated part of the model was identical to the count part of the model with regards to the fixed and random effects' structure. It was investigated whether the model fitted better with a Poisson or a negative binomial distribution using the function anova() and setting the argument "test" to "FALSE". The AIC was lower and the log-likelihood higher for the model with the negative binomial distribution (Poisson Model: AIC = 2020.27, log.Lik = 983.14; Negative Binomial Model: AIC = 1780.82, log.Lik = 862.41). Therefore, the negative binomial distribution was chosen. Then two versions of the same model were compared with either Subject or Group as the random effect using the same strategy as for the error distribution. Adding both factors as random effects would have made the model too complex to run in a reasonable amount of time. The AIC was lower and the log-likelihood higher for the model with Subject as the random effect (model with Subject: AIC = 1780.82, log.Lik = 862.41; model with Group: AIC = 1910.51, log.Lik = 927.25). Therefore, the model with Subject as the random effect was chosen for further inference.

The p-values were extracted from the output of the model summary and corrected by using the conservative Bonferroni Correction Method by setting the argument "method" to "Bonferroni" within the p.adjust() function.

Interobserver reliability was excellent (ICC = 0.875, $p < 0.001$, $n_{Observations} = 18$, $n_{Raters} = 2$)

1.3.2.2 Proportion of Successful Lifts out of Attempts

To investigate how successful the pigs were in relation to how often they attempted to lift the log(s), and to determine the effect that Day, Batch and Task had on this proportion, a GLMM was fitted using the glmer() function in the "lme4" package (version 1.1.32, Bates et al. 2015). The response variable was the proportion of successful lifts out of attempts, which was calculated in the model using the "cbind" function with the number of successful lifts and all attempts to lift. The error distribution was set to binomial. The fixed effects were Day, Batch and Task. The model included a three-way interaction between the fixed effects. The factors Subject and Group were added as random intercept effects. An observation level random effect was only added as random intercept. Day and Task were included as random slopes within the random effects of Subject and Group. An interaction was included between Day and Task within the random effect of Group. The model also included a control option, which was specified before fitting the model using the glmerControl() function. The "optimizer" argument was set to "bobyqa" and the maximum number of function evaluations was set to 100,000.

The correlations among the random slopes and random intercepts were close to 1 or -1 indicating that they were unidentifiable. Therefore, those were removed. Additionally, the estimates and standard deviations for the interactions were very large (Estimates ~10, Standard deviations ~ 1000). One reason as to why this problem was encountered could be that the observations of successful lifts are very low (e.g.: Batch A and Batch B in the task JLL). Another reason could be complete separation or perfect prediction, as some of the levels of our factors were always zero (no successful lifts) and this led to them predicting the response perfectly. Due to the extreme estimates, yet a different model was fitted, in which all interactions between the fixed effects were removed. The problem with the possible complete separation remained, which is why the extracted p-values are to be treated with caution. A full-null model comparison was conducted, with the model without the interactions as the full model, using the function anova() with the argument "test" set to "Chisq", removing the fixed factors from the model, but keeping the rest of the model the same for the null model. The p-values were then extracted from this final model using the drop1() function with the "test" argument set to "Chisq".

2 Results

2.1 Engagement

2.1.1 Interacting with Front of the Box

The full model was a significantly better fit than the null model ($\chi^2 = 30.3406$, df = 4, $p < 0.001$; Fig. 7). The proportion of time spent interacting with the front of the box was significantly lower for batches A and B compared to Batch C in both the JLL task and the ILL task (JLL task: $p_{BatchA} < 0.001$, $p_{BatchB} < 0.001$, Table 3; ILL task: $p_{BatchA} < 0.001$, $p_{BatchB} < 0.001$).

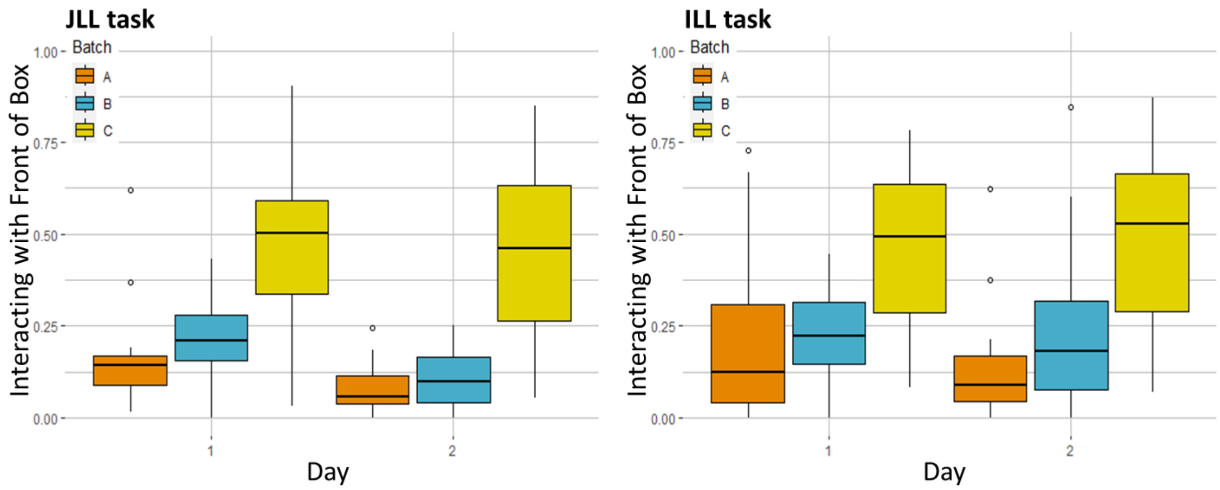

Figure 7: *Boxplots depicting the proportion of total time spent interacting with the front of the JLL box (left) and the ILL box (right), for Batch A (April 2022, orange), Batch B (June 2022, blue) and Batch C (July 2021, yellow) on Day 1 and Day 2.*

2.2 Persistence

2.2.1 Attempt to Lift

The corrected p-values revealed that the number of attempts to lift the log in batches A and B in the JLL task was significantly lower than the number of attempts to lift in Batch C ($p_{BatchA} < 0.001$, $p_{BatchB} < 0.001$; Fig. 8). No significant difference was found between the number of attempts to lift the logs in the ILL task between the three batches ($p_{BatchA} = 0.309$, $p_{BatchB} = 1$).

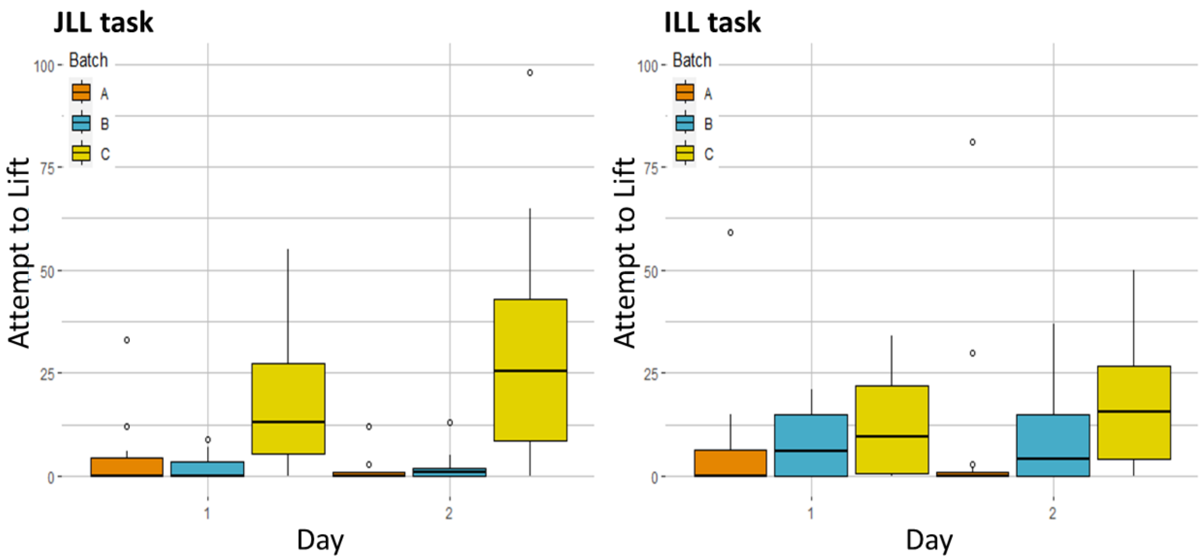

Figure 8: *Boxplots depicting the total number of attempts to lift the log(s) of the JLL box (left) and the ILL box (right), for Batch A (April 2022, orange), Batch B (June 2022, blue) and Batch C (July 2021, yellow) during Day 1 and Day 2.*

2.2.2 Successful Lifts as a Proportion of the Total Number of Attempts

The full model was a significantly better fit than the null model ($\chi^2 = 23.68064$, df = 4, $p < 0.001$; Fig. 9). The drop1() function revealed no significant effect of batch ($p = 0.0703$), but a significant effect of task ($p < 0.001$). However, as previously stated, the model output should be treated with caution, as possible complete separation was encountered when fitting the model. Visual inspection of the plot indicates that the proportion of successful lifts was substantially lower in the batches A and B than in Batch C during the JLL task. Furthermore, when trying to solve the ILL task, Batch A seems to have had a considerably lower proportion of successful lifts than both Batch B and Batch C.

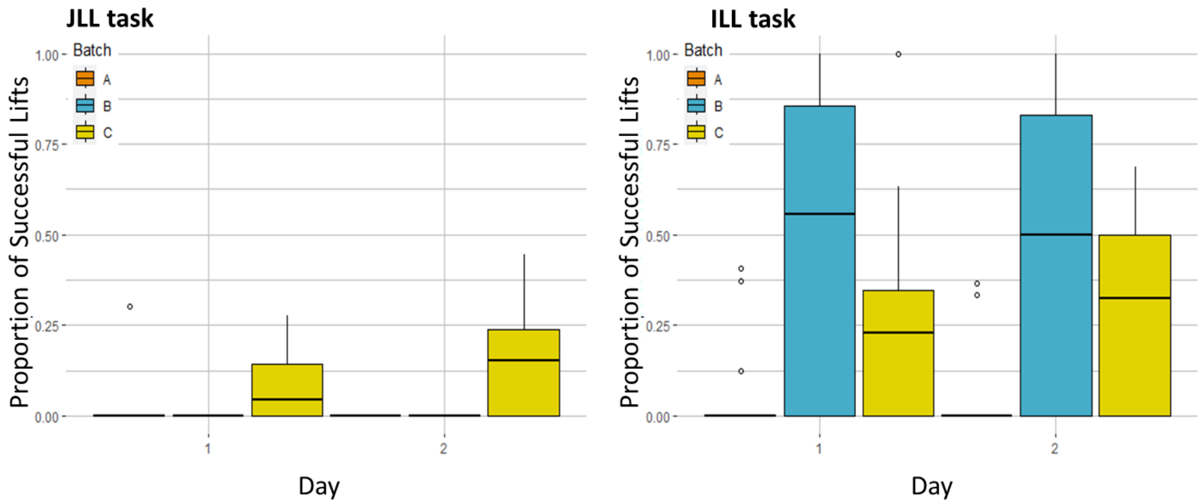

Figure 9: *Boxplots depicting the proportion of successful lifts out of the number of attempts to lift the log(s) of the JLL box (left) and the ILL box (right), for Batch A (April 2022, orange), Batch B (June 2022, blue) and Batch C (July 2021, yellow) during Day 1 and Day 2.*

General Discussion

Pigs have been shown to be highly social animals, are known to participate in cooperative behaviours and previous studies, including the one, which was attempted to be replicated, suggested that pigs are capable of cooperatively solving the JLL task. Even so, across two batches it could not be shown in this study that pigs are always able to reliably solve this specific task together without training. Batches A and B were found to be significantly less successful in solving the JLL task than Batch C. Failing to replicate the previous finding could suggest that the JLL task is not a reliable and robust paradigm to study cooperation in pigs. The results of this study highlight the importance of replication.

Replication is one of the core principles of science, as it is the method by which scientific findings can be verified and gain credibility. Replication studies aim to control for factors such as sampling errors, artifacts or fraud and aids to distinguish between isolated incidents and more generalizable phenomena that can be reproduced by other experimenters or in varying contexts, such as location, methods or time (Schmidt 2009; Frias-Navarro et al. 2020). Even though the consensus among the scientific community is generally high, that replication is an important part of research (Farrar et al. 2021), there seems to be an understanding that new research with significant results is oftentimes the preferred material to publish (Baker 2016). Even so, studies focusing on reproducibility are gaining interest. The majority of researchers from different fields of science reported unsuccessful replication attempts (Baker 2016). Scientists in psychology and related fields attempted to reproduce previously published papers, but only in around 60% of the studies, the formerly significant results could be successfully replicated (Farrar et al. 2020). This finding exemplifies what is referred to as the replication crisis. Studies investigating factors which could influence reproducibility mostly focus on methodological and statistical issues diminishing replication success and the options for improving the documentation and design of methods as well as the statistical analysis to heighten replicability. (Farrar et al. 2020; Farrar et al. 2021; Frias-Navarro et al. 2020; Mansell and Huddy 2018).

In the case of this study, the methodology of the study by McGetrick and colleagues (*in prep.*) was explicitly copied as far as feasible. One difference though, is that experimenters of this study were not the same as in the study which was attempted to be replicated. It has been shown that experimenters can involuntarily influence the outcome of a study, either by their sheer physiology or by unconscious changes in behaviour based on their own beliefs and biases (Rosenthal 2010; Lit et al. 2011; Rosenthal and Fode 1963; Sorge et al. 2014). For example, sex of the experimenter has been shown to influence behavioural and physiological responses in mice (Sorge et al. 2014) but given that all experimenters working with the pigs were female for both studies, this factor can be excluded. However, even though all experimenters were of the same sex, they could have differed in their expectations of the outcome of the study and therefore subconsciously influenced the behaviour of the animals. One possible phenomenon at play could have been the Pygmalion-Effect. This effect describes instances in which the beliefs of an experimenter lead to their expectations being fulfilled due to subtle differences in their behaviour towards the subjects of the study (Rosenthal 2010). For example, Lit and colleagues (2011) found that handler belief could influence the alerting behaviour of trained scent detection dogs, leading to false positive alerts in areas the handlers thought scent should be detectable. More to the point, research in rats has shown that the learning process to solve a maze was negatively influenced when experimenters were led to believe that their testing subjects were dull animals (Rosenthal and Fode 1963). It is possible therefore, that the experimenter in this current study did not believe that the animals would be able to solve the given problem, while the experimenters in the replicated study were of the opposite belief. However, given that the experimenter of this study was aware of the high success of the pigs in the replicated study, and the differences in success were observed already in the first few days of testing, the Pygmalion-Effect seems to be an unlikely explanation. Still, the identity of the experimenter has been shown to affect the outcome of various studies, leading the researchers to speculate that the way the testing subjects were handled, the level of experience of the experimenters or distinct differences in their body odour could have influenced the behaviour of the studied animals (Wahlsten et al. 2003; Bohlen et al. 2014; Chesler et al. 2002). Therefore, experimenter effect remains an elusive factor, and without explicitly testing and controlling for it, it cannot be ruled out that the difference in experimenter had some effect on the behaviour and consequent success of the different batches.

Not all factors, which could impact replication success, are as easily recognisable as differences in experimenters. Especially in studies working with live animals, methodology may seem to be identical in two studies, but other factors, which are in many cases difficult to control for, can additionally influence the outcome and success of a study. One of these factors is called the batch-effect, which describes the fluctuation of different parameters between batches of animals, including personality and interpersonal behaviour (van Leeuwen et al. 2021; O'Bryant et al. 2011; Fuiman et al. 2005).

This batch-effect might have influenced the pigs' success at and behaviour towards the JLL and ILL task. The less successful batches, A and B, were found to show less engagement with both given tasks than Batch C. Batch C showing more interactions with the two boxes, could indicate that the pigs of Batch C were less afraid of and more motivated to interact with the two boxes than the pigs of batches A and B. This could be indicative of a difference in prevailing coping style between the batches. Boldness, more explorative and aggressive behaviour, and higher willingness to participate in a new task is often found in animals with a proactive coping style, while animals with a reactive coping style are often classified as shyer, less curious and less engaged (Prunier et al. 2020; Dugatkin and Alfieri 2003; Scheid and Noë 2010; Finkemeier et al. 2018; Koolhaas et al. 1999). Bolder animals and animals that can be classified as being of the proactive coping style have been shown to be quicker and more successful in learning new tasks, which would point to Batch C as consisting of more pigs with proactive coping style, while the pigs in Batch A and Batch B could have more of a reactive coping style (Svartberg 2002; Range et al. 2006). This notion is supported by the fact that the batches A and B were found to be generally less successful and also less persistent at the JLL task, i.e. they attempted to lift the log less often, than the successful Batch C. More explorative and bold animals have been found to be more persistent, which would underpin the hypothesis that the batches differed in coping style (Massen et al. 2013; Svartberg 2002).

Surprisingly, when looking at the proportion of successful lifts out of the attempts to lift in the ILL task, Batch B appeared to have high rates of success. No definitive significance values could be extracted for the proportion of successful lifts, most likely due to perfect prediction or complete separation of the model. However, when looking at the graphical depiction (Fig. 9) Batch B seems to have at least a similar to, if not even a higher proportion of successful lifts at the ILL task than Batch C. At first glance, this would contradict the previous statement about Batch B consisting of a majority of pigs of the reactive coping style. However, animals with a reactive coping style, even though they are less likely to approach and engage with a novel object or task, have been suggested to be faster in solving a task than proactive animals, if they eventually did approach a novel task (Cole and Quinn 2012). Therefore, the pigs in Batch B might have been more reluctant to approach the task, but after they did approach it, were highly successful in solving it.

Coping style and the consequent behaviours in animals can be influenced by variations in the genetic make-up of animals (Reiner et al. 2009; Koolhaas et al. 1999; Adamczyk et al. 2013; Guo et al. 2015). Therefore, the observed differences in behaviour could be the result of differences in individual experiences of the sows during the pregnancy, as it has been shown that inducing stress in pregnant sows can influence the uterine nutritional and hormonal balance. This can in turn affect the epigenetic make-up of the pigs (Rutherford et al. 2012; Otten et al. 2015), and through this the behaviour and coping style of the pigs as stated above. However, since further testing on these specific pigs is not possible, no definitive statements can be made regarding the genetic differences.

Contrary to this suggestion, that the observed differences in coping styles might be indicative of a difference in individual experiences of the mother sows, studies have found that piglets from the same sow can differ in their coping style (Ruis et al. 2000; Hessing et al. 1993). This could mean, that these results simply indicate that the distribution of coping styles of the pigs, i.e. which coping style was predominant in which batch, was skewed between the batches by sampling luck. However, given the extreme differences in success between the three batches, this explanation is improbable.

Interestingly, the variation in persistence found at the JLL task could not be shown at the ILL task, which could be explained by differences in the social relationships between the batches A and B and those in Batch C. Studies have shown, that social relationship can influence the success of cooperation and the willingness of animals to participate in a given cooperative task (Dale et al. 2020; Massen et al. 2015; Molesti and Majolo 2016; Schwing et al. 2016). Furthermore, it has been shown that social affiliation between animals can influence the level of tolerance of feeding next to each other (Dale et al. 2017). Therefore, the co-feeding tolerance of the pigs in batches A and B might have additionally been lower than in Batch C, which would have made solving the JLL task more difficult, given that two pigs need to work on the task at the same time.

These explanations would substantiate the notion of a measurable batch-effect. However, even though batch-effect has been shown to have an impact on behaviour and learning abilities, the underlying causes for this effect are still under debate (O'Bryant et al. 2011; Fuiman et al. 2005).

An alternative explanation for the observed difference in attempts to lift the logs between the two tasks in Batch A and B, as well as for the differences in overall success at the two tasks, could be that the reward was simply more easily obtainable in the ILL task than in the JLL task. Given that the pigs could solve the ILL task without a partner, it is likely that they got to the apple pieces sooner, after fewer attempts than when attempting to solve the JLL task. Reaching the reward could have acted as an additional motivator and could have kept the pigs trying to solve the ILL task, as opposed to the JLL task. Following the same line of argumentation, social facilitation could have also played a part in this difference between attempts in the two tasks. Studies have shown that observing conspecifics consuming a food reward can elicit consummatory behaviour in the observing animal (Hsia and Wood-Gush 1984; Zentall and Levine 1972; Keeling and Hurnik 1993). This phenomenon could have influenced the behaviour of the pigs during the ILL task. When some pigs solved the ILL task, and reached the reward, others could have been motivated to try again in order to obtain a reward themselves. When given the opportunity to attempt to solve the JLL task, this would be unlikely to happen, as one pig can never succeed alone. Therefore, the possibility of a few pigs solving the JLL task by trial and error is lower than for one determined pig to solve the ILL task. This notion is substantiated by the observable learning curve in the ILL task of Batch B, as well as of Batch A (Fig. 6).

Surprisingly, even though both Batch A and Batch B had an observable learning curve, Batch B was significantly more successful at the ILL task than Batch A, with a similar success rate to Batch C. This could mean that the coping style did not only differ between the two batches of this study and Batch C, but also between Batch A and B. Even though coping style is oftentimes described and thought of as bimodal, i.e. proactive vs. reactive, some research suggests, that individual coping style is more complex, actually being a continuous variable, with individuals falling onto a certain part of the spectrum (reviewed in Koolhaas et al. 1999). This could suggest that in one batch the majority of the individuals fell into a similar range of this spectrum, with this range differing between all three batches.

An alternative explanation for the differences in success at the ILL task between Batch A and Batch B, could be the fact that the learning phase of these batches took place at different times of the year. While Batch A was studied in April, Batch B was studied in June. As the temperatures were still lower in April, and the heated sleeping area was on the opposite end of the enclosures to the box, it could be that the pigs were less motivated in April to leave the sleeping area and interact with the box. In June on the other hand, the temperatures were higher, and the coolest place of the enclosures was directly above the slatted floor area, which was the same area in which the box was placed. Therefore, Batch B might have been more successful at the ILL task than Batch A, simply because of time of year. This notion is substantiated by the fact, that the highly successful Batch C was studied at a similar time of year to Batch B, namely in July of 2021.

The marked difference in success at the ILL task between the batches A and B is in stark contrast to the fact that both batches performed significantly less successfully in the JLL task than Batch C. This could mean that the experience with the ILL task does not transfer well to the JLL task as one would expect given that the movement needed to be successful is identical for both tasks, save for the coordination with a partner. However, the success at the ILL task of Batch B could still have acted as a motivator, prompting the pigs to try to solve the JLL task, as there is an evident increase in successful lifts for Batch B in the JLL task.

In conflict with this line of argument is the aforementioned observation that Batch B had significantly less success at the JLL task and showed significantly less engagement with both boxes than Batch C. Therefore, the success at the ILL task does not seem to keep the pigs sufficiently motivated to interact with and solve the JLL task as well, and it suggests that there are other factors at play that are difficult to pinpoint from the data we have gathered so far.

In conclusion, the differences in success as well as the observed differences in engagement and persistence, especially at the JLL task, could be indicative of differences in the prevalence of coping styles and the nature of the social relationships between the three batches. These differences substantiate the notion of a measurable effect of batch on cooperative behaviour. Even though a single replication study cannot invalidate previous findings (Farrar et al. 2020), failing to replicate the findings of the study by McGetrick et al. (*in prep.*) highlights the importance of replications and reviews especially in the behavioural and cognitive sciences.

The observed batch-effect nicely demonstrates how unreliable single studies on behaviour and cognition can be. Despite standardisation being viewed as the key for reproducibility, geno- and especially phenotypic plasticity makes it difficult for research on live animals to be completely standardised across multiple studies. Researchers are proposing to introduce controlled variability into studies, which is called "systematic heterogenization", in order to increase external validity and reproducibility (Kortzfleisch et al. 2022; Richter et al. 2010). However, to date no clear guidelines on how to implement this approach in single studies have been formulated. Even so, research suggests that the introduction of small changes such as variations in time of testing, as well as multiple small batch testing increases confidence in the results, making them more robust against variability between different experiments and thereby increasing reproducibility (Bodden et al. 2019; Karp et al. 2020). This accentuates the implication of the results of the current study, that the outcome of single, highly standardised studies in behavioural research can be highly unreliable and emphasises the need for further insights into factors influencing the success of pigs at the JLL task, before a definitive statement can be made about the robustness of this paradigm.

References

Adamczyk, K.; Pokorska, J.; Makulska, J.; Earley, B.; Mazurek, M. (2013): Genetic analysis and evaluation of behavioural traits in cattle. In *Livestock Science* 154 (1-3), pp. 1–12. DOI: 10.1016/j.livsci.2013.01.016.

Baker, M. (2016): Is There A Reproducibility Crisis? A Nature survey lifts the lid on how researcher view the 'crisis' rocking science and what theiy think will help. In *Nature* (533), 452-454. DOI: 10.1126/science.aac4716.

Bates, D.; Mächler, M.; Bolker, B.; Walker, S. (2015): Fitting Linear Mixed-Effects Models Using lme4. In *Journal of Statistical Software* 67 (1). DOI: 10.18637/jss.v067.i01.

Bodden, Carina; Kortzfleisch, Vanessa Tabea von; Karwinkel, Fabian; Kaiser, Sylvia; Sachser, Norbert; Richter, S. Helene (2019): Heterogenising study samples across testing time improves reproducibility of behavioural data. In *Scientific reports* 9 (1), p. 8247. DOI: 10.1038/s41598-019-44705-2.

Bohlen, M.; Hayes, E.; Bohlen, B.; Bailoo, J.; Crabbe, J.; Wahlsten, D. (2014): Experimenter effects on behavioral test scores of eight inbred mouse strains under the influence of ethanol. In *Behavioural brain research* 272, pp. 46–54. DOI: 10.1016/j.bbr.2014.06.017.

Brooks, M.; Kristensen, K.; van Benthem, K.; Magnusson, A.; Berg, C.; Nielsen, A. et al. (2017): glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. In *The R Journal* 9 (2), p. 378. DOI: 10.32614/RJ-2017-066.

Byrd, C.; Johnson, J.; Radcliffe, J.; Craig, B.; Eicher, S.; Lay, D. (2020): Nonlinear analysis of heart rate variability for evaluating the growing pig stress response to an acute heat episode. In *Animal: an international journal of animal bioscience* 14 (2), pp. 379–387. DOI: 10.1017/S1751731119001630.

Chesler, Elissa J.; Wilson, Sonya G.; Lariviere, William R.; Rodriguez-Zas, Sandra L.; Mogil, Jeffrey S. (2002): Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. In *Neuroscience and Biobehavioral Reviews* 26 (8), pp. 907–923. DOI: 10.1016/s0149-7634(02)00103-3.

Clutton-Brock, T.; Russell, A.; Sharpe, L. (2004): Behavioural tactics of breeders in cooperative meerkats. In *Animal Behaviour* 68 (5), pp. 1029–1040. DOI: 10.1016/j.anbehav.2003.10.024.

Cockburn, A. (2006): Prevalence of different modes of parental care in birds. In *Proceedings. Biological sciences* 273 (1592), pp. 1375–1383. DOI: 10.1098/rspb.2005.3458.

Cole, E.; Quinn, J. (2012): Personality and problem-solving performance explain competitive ability in the wild. In *Proceedings. Biological sciences* 279 (1731), pp. 1168–1175. DOI: 10.1098/rspb.2011.1539.

Czaczkes, T.; Ratnieks, F. (2013): Cooperative transport in ants (Hymenoptera: Formicidae) and elsewhere. In *Myrmecological News* (18), pp. 1–11.

Dale, R.; Marshall-Pescini, S.; Range, F. (2020): What matters for cooperation? The importance of social relationship over cognition. In *Scientific reports* 10 (1), p. 11778. DOI: 10.1038/s41598-020-68734-4.

Dale, R.; Range, F.; Stott, L.; Kotrschal, K.; Marshall-Pescini, S. (2017): The influence of social relationship on food tolerance in wolves and dogs. In *Behav Ecol Sociobiol* 71 (7), p. 107. DOI: 10.1007/s00265-017-2339-8.

Dugatkin, L. (1997): The Evolution of Cooperation. Four paths to the evolution and maintenance of cooperative behavior. In *BioScience*, pp. 355–362.

Dugatkin, L. (2002): Animal cooperation among unrelated individuals. In *Die Naturwissenschaften* 89 (12), pp. 533–541. DOI: 10.1007/s00114-002-0379-y.

Dugatkin, L.; Alfieri, M. (2003): Boldness, behavioral inhibition and learning. In *Ethology Ecology & Evolution* 15 (1), pp. 43–49. DOI: 10.1080/08927014.2003.9522689.

Farrar, B.; Boeckle, M.; Clayton, N. (2020): Replications in Comparative Cognition: What Should We Expect and How Can We Improve? In *Animal behavior and cognition* 7 (1), pp. 1–22. DOI: 10.26451/abc.07.01.02.2020.

Farrar, B.; Ostojic, L.; Clayton, N. (2021): The hidden side of animal cognition research: Scientists' attitudes toward bias, replicability and scientific practice. In *PLOS ONE*.

Fels, M.; Hoy, S.; Hartung, J. (2012): Influence of origin litter on social rank, agonistic behaviour and growth performance of piglets after weaning. In *Applied Animal Behaviour Science* 139 (3-4), pp. 225–232. DOI: 10.1016/j.applanim.2012.03.017.

Finkemeier, M.; Langbein, J.; Puppe, B. (2018): Personality Research in Mammalian Farm Animals: Concepts, Measures, and Relationship to Welfare. In *Frontiers in veterinary science* 5, p. 131. DOI: 10.3389/fvets.2018.00131.

Focardi, S.; Morimando, F.; Capriotti, S.; Ahmed, A.; Genov, P. (2015): Cooperation improves the access of wild boars (Sus scrofa) to food sources. In *Behavioural processes* 121, pp. 80–86. DOI: 10.1016/j.beproc.2015.10.019.

Forstmeier, W.; Schielzeth, H. (2011): Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. In *Behav Ecol Sociobiol* 65 (1), pp. 47–55. DOI: 10.1007/s00265-010-1038-5.

Fraser, D.; Kramer, D.; Pajor, E.; Weary, D. (1995): Conflict and cooperation: sociobiological principles and the behaviour of pigs. In *Applied Animal Behaviour Science* (44), pp. 139–157.

Friard, O.; Gamba, M. (2016): BORIS : a free, versatile open-source event-logging software for video/audio coding and live observations. In *Methods Ecol Evol* 7 (11), pp. 1325–1330. DOI: 10.1111/2041-210X.12584.

Frias-Navarro, D.; Pascual-Llobell, J.; Pascual-Soler, M.; Perezgonzalez, J.; Berrios-Riquelme, J. (2020): Replication crisis or an opportunity to improve scientific production? In *Eur. J. Educ.* 55 (4), pp. 618–631. DOI: 10.1111/ejed.12417.

Fuiman, L.; Cowan, J.; Smith, M.; O'Neal, J. (2005): Behavior and recruitment success in fish larvae: variation with growth rate and the batch effect. In *Canadian Journal of Fisheries and Aquatic Sciences* 62 (6), pp. 1337–1349. DOI: 10.1139/f05-053.

Gonyou, H. (2001): The social behaviour of pigs. In : Social behaviour in farm animals. 1st ed. UK: CABI Publishing, pp. 147–176.

Guo, Y.; Zhang, Z.; Ma, J.; Ai, H.; Ren, J.; Huang, L. (2015): A genomewide association study of feed efficiency and feeding behaviors at two fattening stages in a White Duroc × Erhualian F2 population1. In *Journal of Animal Science*.

Hatchwell, B. (2009): The evolution of cooperative breeding in birds: kinship, dispersal and life history. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* (364), pp. 3217–3227.

Hessing, M.J.C.; Hagelsø, A. M.; van Beek, J.A.M.; Wiepkema, R. P.; Schouten, W.G.P.; Krukow, R. (1993): Individual behavioural characteristics in pigs. In *Applied Animal Behaviour Science* 37 (4), pp. 285–295. DOI: 10.1016/0168-1591(93)90118-9.

Hirata, S.; Fuwa, K. (2007): Chimpanzees (Pan troglodytes) learn to act with other individuals in a cooperative task. In *Primates* 48 (1), pp. 13–21. DOI: 10.1007/s10329-006-0022-1.

Hsia, L.; Wood-Gush, D. (1984): Social facilitation in the feeding behaviour of pigs and the effect of rank. In *Applied Animal Ethology* 11 (3), pp. 265–270. DOI: 10.1016/0304-3762(84)90033-6.

Jacobs, I.; Osvath, M. (2015): The string-pulling paradigm in comparative psychology. In *Journal of comparative psychology (Washington, D.C. : 1983)* 129 (2), pp. 89–120. DOI: 10.1037/a0038746.

Karp, Natasha A.; Wilson, Zena; Stalker, Eve; Mooney, Lorraine; Lazic, Stanley E.; Zhang, Bairu; Hardaker, Elizabeth (2020): A multi-batch design to deliver robust estimates of efficacy and reduce animal use - a syngeneic tumour case study. In *Scientific reports* 10 (1), p. 6178. DOI: 10.1038/s41598-020-62509-7.

Keeling, L.; Hurnik, J. (1993): Chickens show socially facilitated feeding behaviour in response to a video image of a conspecific. In *Applied Animal Behaviour Science* 36 (2-3), pp. 223–231. DOI: 10.1016/0168-1591(93)90012-E.

Kennedy, R.; Evans, C.; McDonald, P. (2009): Individual distinctiveness in the mobbing call of a cooperative bird, the noisy miner Manorina melanocephala. In *J Avian Biol* 40 (5), pp. 481–490. DOI: 10.1111/j.1600-048X.2008.04682.x.

Koglmüller, M.; Wondrak, M.; Camerlink, I.; Rault, J.; Huber, L. (2021): Are free-ranging Kune Kune pigs (Sus scrofa domesticus) able to solve a cooperative task? In *Applied Animal Behaviour Science* 240, p. 105340. DOI: 10.1016/j.applanim.2021.105340.

Koolhaas, J.; Korte, S.; Boer, S. de; van der Vegt, B.; van Reenen, C.; Hopster, H. et al. (1999): Coping styles in animals: current status in behavior and stress-physiology. In *Neuroscience and Biobehavioral Reviews* 23 (7), pp. 925–935. DOI: 10.1016/S0149-7634(99)00026-3.

Kortzfleisch, Vanessa Tabea von; Ambrée, Oliver; Karp, Natasha A.; Meyer, Neele; Novak, Janja; Palme, Rupert et al. (2022): Correction: Do multiple experimenters improve the reproducibility of animal studies? In *PLoS biology* 20 (12), e3001948. DOI: 10.1371/journal.pbio.3001948.

Lit, L.; Schweitzer, J.; Oberbauer, A. (2011): Handler beliefs affect scent detection dog outcomes. In *Anim Cogn* 14 (3), pp. 387–394. DOI: 10.1007/s10071-010-0373-2.

Lönnstedt, O.; Ferrari, M.; Chivers, D. (2014): Lionfish predators use flared fin displays to initiate cooperative hunting. In *Biology letters* 10 (6). DOI: 10.1098/rsbl.2014.0281.

Mansell, W.; Huddy, V. (2018): The Assessment and Modeling of Perceptual Control: A Transformation in Research Methodology to Address the Replication Crisis. In *Review of General Psychology* 22 (3), pp. 305–320. DOI: 10.1037/gpr0000147.

Marler, P. (1956): Behaviour of the chaffinch Fringilla coelebs. Available online at https://www.jstor.org/stable/30039131.

Marshall-Pescini, S.; Schwarz, J.; Kostelnik, I.; Virányi, Z.; Range, F. (2017): Importance of a species' socioecology: Wolves outperform dogs in a conspecific cooperation task. In *Proceedings of the National Academy of Sciences of the United States of America* 114 (44), pp. 11793–11798. DOI: 10.1073/pnas.1709027114.

Massen, J.; Antonides, A.; Arnold, A.; Bionda, T.; Koski, S. (2013): A behavioral view on chimpanzee personality: exploration tendency, persistence, boldness, and tool-orientation measured with group experiments. In *American journal of primatology* 75 (9), pp. 947–958. DOI: 10.1002/ajp.22159.

Massen, J.; Behrens, F.; Martin, J.; Stocker, M.; Brosnan, S. (2019): A comparative approach to affect and cooperation. In *Neuroscience and Biobehavioral Reviews* (107), pp. 370–387.

Massen, J.; Ritter, C.; Bugnyar, T. (2015): Tolerance and reward equity predict cooperation in ravens (Corvus corax). In *Scientific reports* 5, p. 15021. DOI: 10.1038/srep15021.

McGetrick, J.; Brosche, K.; Nanchen, C.; Rault, J. (in prep.): Lack of evidence that pigs understand the need for a partner in a cooperative task.

Molesti, S.; Majolo, B. (2016): Cooperation in wild Barbary macaques: factors affecting free partner choice. In *Anim Cogn* 19 (1), pp. 133–146. DOI: 10.1007/s10071-015-0919-4.

Muro, C.; Escobedo, R.; Spector, L.; Coppinger, R. (2011): Wolf-pack (Canis lupus) hunting strategies emerge from simple rules in computational simulations. In *Behavioural processes* 88 (3), pp. 192–197. DOI: 10.1016/j.beproc.2011.09.006.

Noë, R. (2006): Cooperation experiments: coordination through communication versus acting apart together. In *Animal Behaviour* 71 (1), pp. 1–18. DOI: 10.1016/j.anbehav.2005.03.037.

Nolen, M.; Lucas, J. (2009): Asymmetries in mobbing behaviour and correlated intensity during predator mobbing by nuthatches, chickadees and titmice. In *Animal Behaviour* 77 (5), pp. 1137–1146. DOI: 10.1016/j.anbehav.2009.01.023.

O'Bryant, A.; Allred, R.; Maldonado, M.; Cormack, L.; Jones, T. (2011): Breeder and batch-dependent variability in the acquisition and performance of a motor skill in adult Long-Evans rats. In *Behavioural brain research* 224 (1), pp. 112–120. DOI: 10.1016/j.bbr.2011.05.028.

Otten, W.; Kanitz, E.; Tuchscherer, M. (2015): The impact of pre-natal stress on offspring development in pigs. In *The Journal of Agricultural Science* 153 (5), pp. 907–919. DOI: 10.1017/S0021859614001361.

Péron, F.; Rat-Fischer, L.; Lalot, M.; Nagle, L.; Bovet, D. (2011): Cooperative problem solving in African grey parrots (Psittacus erithacus). In *Anim Cogn* 14 (4), pp. 545–553. DOI: 10.1007/s10071-011-0389-2.

Plotnik, J.; Lair, R.; Suphachoksahakun, W.; Waal, F. de (2011): Elephants know when they need a helping trunk in a cooperative task. In *Proceedings of the National Academy of*

Sciences of the United States of America 108 (12), pp. 5116–5121. DOI: 10.1073/pnas.1101765108.

Prunier, A.; Averos, X.; Dimitrov, I.; Edwards, S.; Hillmann, E.; Holinger, M. et al. (2020): Review: Early life predisposing factors for biting in pigs. In *Animal : an international journal of animal bioscience* 14 (3), pp. 570–587. DOI: 10.1017/S1751731119001940.

R Core Team (2023): R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available online at https://www.R-project.org/.

Range, F.; Bugnyar, T.; Schlögl, C.; Kotrschal, K. (2006): Individual and sex differences in learning abilities of ravens. In *Behavioural processes* 73 (1), pp. 100–106. DOI: 10.1016/j.beproc.2006.04.002.

Rault, J. (2019): Be kind to others: Prosocial behaviours and their implications for animal welfare. In *Applied Animal Behaviour Science* 210, pp. 113–123. DOI: 10.1016/j.applanim.2018.10.015.

Rault, J.; Camerlink, I.; Goumon, S.; Mundry, R.; Špinka, M. (2021): The Joint Log-Lift Task: A Social Foraging Paradigm. In *Frontiers in veterinary science* 8, p. 745627. DOI: 10.3389/fvets.2021.745627.

Rault, J.; Camerlink, I.; Goumon, S.; Špinka, M. (2020): Development of the joint-lift task in pigs: hints on cooperation? In *ISAE 2020 Global Virtual Meeting: Online Programme Book*.

Reiner, G.; Köhler, F.; Berge, T.; Fischer, R.; Hübner-Weitz, K.; Scholl, J.; Willems, H. (2009): Mapping of quantitative trait loci affecting behaviour in swine. In *Animal genetics* 40 (4), pp. 366–376. DOI: 10.1111/j.1365-2052.2008.01847.x.

Richter, S. Helene; Garner, Joseph P.; Auer, Corinna; Kunert, Joachim; Würbel, Hanno (2010): Systematic variation improves reproducibility of animal experiments. In *Nature methods* 7 (3), pp. 167–168. DOI: 10.1038/nmeth0310-167.

Rizopoulos, D. (2023): GLMMadaptive: Generalized Linear Mixed Models using Adaptive Gaussian. Version R package version 0.8-8. Available online at https://CRAN.R-project.org/package=GLMMadaptive.

Rodrigues, A.; Barker, J.; Robinson, E. (2022): From inter-group conflict to inter-group cooperation: insights from social insects. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 377 (1851), p. 20210466. DOI: 10.1098/rstb.2021.0466.

Rosenthal, R. (2010): Pygmalion Effect. In I. Weiner, W. Craighead (Eds.): The Corsini Encyclopedia of Psychology: Wiley, pp. 1–2.

Rosenthal, R.; Fode, K. (1963): The effect of experimenter bias on the performance of the albino rat. In *Syst. Res.* 8 (3), pp. 183–189. DOI: 10.1002/bs.3830080302.

Ruis, Marko A.W; te Brake, Joop H.A; van de Burgwal, Johan A.; Jong, Ingrid C. de; Blokhuis, Harry J.; Koolhaas, Jaap M. (2000): Personalities in female domesticated pigs: behavioural and physiological indications. In *Applied Animal Behaviour Science* 66 (1-2), pp. 31–47. DOI: 10.1016/S0168-1591(99)00070-2.

Rutherford, K.; Donald, R.; Arnott, G.; Rooke, J.; Dixon, L.; Mehers, J. et al. (2012): Farm animal welfare: assessing risks attributable to the prenatal environment. In *Anim. welf.* 21 (3), pp. 419–429. DOI: 10.7120/09627286.21.3.419.

Scheid, C.; Noë, R. (2010): The performance of rooks in a cooperative task depends on their temperament. In *Animal cognition* 13 (3), pp. 545–553. DOI: 10.1007/s10071-009-0305-1.

Schmidt, Stefan (2009): Shall we Really do it Again? The Powerful Concept of Replication is Neglected in the Social Sciences. In *Review of General Psychology* 13 (2), pp. 90–100. DOI: 10.1037/a0015108.

Schwing, R.; Jocteur, E.; Wein, A.; Noë, R.; Massen, J. (2016): Kea cooperate better with sharing affiliates. In *Anim Cogn* 19 (6), pp. 1093–1102. DOI: 10.1007/s10071-016-1017-y.

Schwing, R.; Reuillon, L.; Conrad, M.; Noë, R.; Huber, L. (2020): Paying attention pays off: Kea improve in loose-string cooperation by attending to partner. In *Ethology* 126 (2), pp. 246–256. DOI: 10.1111/eth.12984.

Seed, A.; Clayton, N.; Emery, N. (2008): Cooperative problem solving in rooks (Corvus frugilegus). In *Proceedings. Biological sciences* 275 (1641), pp. 1421–1429. DOI: 10.1098/rspb.2008.0111.

Smith, J.; Swanson, E.; Reed, D.; Holekamp, K. (2012): Evolution of Cooperation among Mammalian Carnivores and Its Relevance to Hominin Evolution. In *Current Anthropology* 53 (S6), S436-S452. DOI: 10.1086/667653.

Sorge, R.; Martin, L.; Isbester, K.; Sotocinal, S.; Rosen, S.; Tuttle, A. et al. (2014): Olfactory exposure to males, including men, causes stress and related analgesia in rodents. In *Nature methods* 11 (6), pp. 629–632. DOI: 10.1038/nmeth.2935.

Stander, P. (1992): Cooperative hunting in lions: the role of the individual. In *Behavioral Ecology and Sociobiology* (29), pp. 445–454.

Svartberg, K. (2002): Shyness–boldness predicts performance in working dogs. In *Applied Animal Behaviour Science* (79), pp. 157–174.

Taborsky, M.; Taborsky, B. (2015): Evolution of genetic and physiological mechanisms of cooperative behaviour. In *Current Opinion in Behavioral Sciences* 6, pp. 132–138. DOI: 10.1016/j.cobeha.2015.11.001.

van Leeuwen, E.; DeTroy, S.; Kaufhold, S.; Dubois, C.; Schütte, S.; Call, J.; Haun, D. (2021): Chimpanzees behave prosocially in a group-specific manner. In *Science Advances* 7 (9). DOI: 10.1126/sciadv.abc7982.

Wahlsten, Douglas; Metten, Pamela; Phillips, Tamara J.; Boehm, Stephen L.; Burkhart-Kasch, Sue; Dorow, Janet et al. (2003): Different data from different labs: lessons from studies of gene-environment interaction. In *Journal of Neurobiology* 54 (1), pp. 283–311. DOI: 10.1002/neu.10173.

West, S.; Griffin, A.; Gardner, A. (2007): Evolutionary explanations for cooperation. In *Current biology : CB* 17 (16), R661-72. DOI: 10.1016/j.cub.2007.06.004.

Wickham, H. (2016): ggplot2. Elegant graphics for data analysis. With assistance of Carson Sievert. Second edition. Switzerland: Springer (Use R!).

Wong, M.; Balshine, S. (2011): The evolution of cooperative breeding in the African cichlid fish, Neolamprologus pulcher. In *Biological reviews of the Cambridge Philosophical Society* 86 (2), pp. 511–530. DOI: 10.1111/j.1469-185X.2010.00158.x.

Yip, E.; Powers, K.; Avilés, L. (2008): Cooperative capture of large prey solves scaling challenge faced by spider societies. In *PNAS* (105), pp. 11818–11822.

Zentall, T.; Levine, J. (1972): Observational learning and social facilitation in the rat. In *Science (New York, N.Y.)* 178 (4066), pp. 1220–1221. DOI: 10.1126/science.178.4066.1220.

Appendix

List of Figures

List of Tables

Table 2*: Full model output including estimate, standard error, z-value, and p-value for model assessing the proportion of total time spent interacting with the front of the JLL box with Batch C as reference.*

Engagement JLL				
Term	Estimate	Standard Error	z-value	p-value
Intercept	-1.000	0.129	-7.747	< 0.001
Day*	-0.122	0.049	-2.469	0.0135
Batch A	-1.387	0.183	-7.573	< 0.001
Batch B	-1.077	0.182	-5.900	< 0.001
Task ILL	0.064	0.129	0.497	0.6192

*z-transformed to mean = 0 and sd = 1.

Table 3*: Full model output including estimate, standard error, z-value, and p-value for model assessing the proportion of total time spent interacting with the front of the ILL box with Batch C as reference.*

. **Engagement ILL**				
Term	Estimate	Standard Error	z-value	p-value
Intercept	-0.936	0.129	-7.263	< 0.001
Day*	-0.122	0.049	-2.469	0.0135
Batch A	-1.387	0.183	-7.573	< 0.001
Batch B	-1.077	0.183	-5.900	< 0.001
Task JLL	-0.064	0.129	-0.497	0.6192

*z-transformed to mean = 0 and sd = 1.

Table 4: *Full model output including estimate, standard error, z-value, and p-value for the model assessing attempts to lift at the JLL box with Batch C as reference.*

Attempt to Lift JLL				
Term	Estimate	Standard Error	z-value	p-value+
Intercept	2.772	0.1872	-7.747	< 0.001
Day*	0.201	0.0690	-2.469	0.0424
Batch A	-2.524	0.183	0.384	< 0.001
Batch B	-2.060	0.182	0.332	< 0.001
Task ILL	-0.344	0.129	0.103	0.01
Day*:BatchB	-0.547	0.203	-2.696	1
Day*:BatchA	-1.064	0.258	-4.114	0.0017
Day*:TaskILL	-0.043	0.102	-0.420	1
BatchB:TaskILL	1.597	0.231	6.913	< 0.001
BatchA:TaskILL	1.763	0.308	5.725	< 0.001
Day*:BatchB:TaskILL	0.433	0.243	1.780	0.9003
Day*:BatchA:TaskILL	0.309	0.319	0.968	1

*z-transformed to mean = 0 and sd = 1.

+p-values presented are corrected using the Bonferroni method

***Table 5**: Full model output including estimate, standard error, z-value, and p-value for the model assessing attempts to lift at the ILL box with Batch C as reference.*

Attempt to Lift ILL				
Term	Estimate	Standard Error	z-value	p-value[+]
Intercept	2.430	0.192	12.625	< 0.001
Day*	0.158	0.076	2.077	0.4533
Batch A	-0.763	0.342	-2.230	0.3091
Batch B	-0.464	0.307	-1.510	1
Task JLL	0.344	0.103	3.342	0.01
Day*:BatchB	-0.114	0.136	-0.837	0.4028
Day*:BatchA	-0.754	0.198	-3.801	< 0.001
Day*:TaskJLL	0.043	0.102	0.423	1
BatchB:TaskJLL	-1.597	0.231	-6.912	< 0.001
BatchA:TaskJLL	-1.764	0.308	-5.727	< 0.001
Day*:BatchB:TaskJLL	-0.435	0.243	-1.787	0.8880
Day*:BatchA:TaskJLL	-0.309	0.320	-0.967	1

*z-transformed to mean = 0 and sd = 1.

[+]p-values presented are corrected using the Bonferroni method

***Table 6**: Full model output including estimate, standard error, z-value and p-value for the model assessing the proportion of successful lifts out of the number of attempts to lift at the JLL box with Batch C as reference.*

Proportion of Successful Lifts JLL				
Term	Estimate	Standard Error	z-value	p-value+
Intercept	-2.251	0.201	-11.199	
Day*	0.122	0.043	2.848	0.0154
Batch A	-0.568	0.256	-2.219	0.0703
Batch B	-0.125	0.318	-0.392	
Task ILL	1.524	0.337	4.516	< 0.001

*z-transformed to mean = 0 and sd = 1.

+p-values presented are from the output of the drop1 function