From the Messerli Research Institute

University of Veterinary Medicine, Vienna

Unit of Comparative Cognition
(Head: Univ.-Prof. Mag. Dr.rer.nat. Ludwig Huber)

Quantity Discrimination in Kea (*Nestor notabilis*)

Master Thesis

Submitted in partial fulfilment of the requirements for the degree of

Master of Science (M.Sc.)

at the University of Veterinary Medicine, Vienna

by

David Zdravko Beraha

Vienna, September 2019

Supervisors:

Univ.-Prof. Mag. Dr.rer.nat. Ludwig Huber

Raoul Schwing, PhD

Reviewer:

Priv.-Doz. Sarah Marshall-Pescini, PhD

Index

1. Introduction

1.1. Quantity Discrimination

Numerical abilities in animals have fascinated researchers and lay people alike and are a topic of special interest in comparative cognitive science to this day (Agrillo & Beran, 2013). While several highly popularised instances of numerical ability displays have turned out to be frauds (e.g. the famous "Clever Hans"), researchers of comparative cognition have established the existence of proto-numerical abilities – meaning abilities that enable to perceive and differentiate amounts but are not explicitly counting – in many species (Drake et al., 2000). One of those proto-numerical abilities is quantity discrimination, the ability to distinguish between different amounts of a given stimulus. It is the ability to recognize the relationship between stimuli based on amount, irrespective of the precise appearance of those stimuli (Moll & Nieder, 2014). It is viewed as the basis for more advanced numerical abilities, and hence offers a stepping-stone for researchers trying to trace the evolutionary development of numerical capabilities in animals (Hauser & Spelke, 2004).

1.2. Theories about quantity discrimination

The property called "numerosity" or "numerousness" is defined as an attribute that can be discriminated when regarding a collection of objects (Stevens, [1938] 1939). It is exhibited in degrees and can be used in order to rank different groupings. When two such groupings are compared and one contains more items than the other it is said to have the greater numerosity. The difference between two sets can be conceptualised in at least two distinct ways. Firstly, the absolute difference, which is the result of the subtraction of the smaller quantity from the larger, and secondly, the relative difference, which results from the division of the larger of the two quantities by the smaller one (i.e., the inverse of the ratio).

Psychologists Davis and Pérusse (1988) distinguish between several types of numerical competence in animals, suggesting that different cognitive or perceptual processes underlie each of these types. They conceptualized categories of relative numerousness judgments, dividing numerical competences according to the labels of "subitizing", "estimation", and "counting". Davis and Pérusse describe subitizing and estimation as the most basic level of decision processes in this regard, because no concept of absolute number is a prerequisite for the animal. It perceives numerical differences on grounds of the order in magnitude, in other words according to if some amount is more or less than another (Davis & Pérusse, 1988; Emmerton, 2001).

Introduction

The term *subitizing* has a long history in human psychological studies of quantification. Initially it referred to the ability to quickly perceive and enumerate quantities in the range from one to seven (Taves, 1941). Subsequently, experiments showed that the reaction time in the range between one and four was on average fast, but increased steadily and linearly if tested beyond that range, while accuracy of correct positives decreased continuously. This indicates that other cognitive mechanisms were being utilized when discriminating higher quantities (Mandler & Shebo, 1982). Nowadays the term subitizing is used in the literature to describe a type of recognizing numerical patterns that is utilized to assess small quantities of items in a short period of time (Anobile et al., 2015). In this regard the idea has been formulated that these small amounts are perceived and mentally equated with a geometric figure as for example a triangle, thus providing the formation of a specific quantity judgement (Dehaene, 1992). According to Davis and Pérusse this can be seen as "perceptual shortcut" by which the process of pattern recognition is used to quickly assign numerical tags up to the amount of five elements (Davis & Pérusse, 1988). When considering animal cognition in regard to experimental tasks, in which the subjects have to decide on picking an amount based on specific parameters, they could choose by using this process without specifically referring to the actual numerical value. As the pattern is perceived equivalent to the amount, it is rendering a possibility for quantity judgements. The existence of such a subitizing process that can be separated from other ways of perceiving quantities is still in question though (Davis & Pérusse, 1988; Pepperberg, 1999).

The process of *estimation* is describing the ability to make judgements about comparably larger amounts of quantities without counting out every single element, but instead making an approximation of their sum. Similar to subitizing this happens in a relatively short amount of time but with larger numbers. Until the beginning of the 1990s this process was mostly attributed to humans but its possible existence in animals has also been acknowledged. There has been a revision of the term in the literature, basically dividing it up into what can be termed "simple estimation" which works in a comparable manner as subitizing and "enhanced estimation" which designates meaningful number tags but without involving the process of counting (Davis & Pérusse, 1988).

When engaging in the mechanism of *counting*, due to a process of enumerating absolute numbers according to the prevalence of items in a set, quantities can be precisely discriminated. For this purpose each element is labelled one after the other and thus assigned an individual tag. The cardinal or absolute number that is applied to a set is equal to the number that is reached when finishing this process of labelling and designates the numerosity of the group, implying the possibility of infinite enumeration. The authors

emphasize that quantity discrimination is not equal to *counting* and argue that the latter is rarely observed amongst non-human animals and probably reserved for humans (Davis & Pérusse, 1988). A basic necessity for counting entails that groupings displaying the same numerosity are being designated the same sign, also referred to as "*numeral*" (Stevens, [1938] 1939). Non-human animals like apes and the African grey parrot Alex have been able to match numerals to cardinal numbers precisely after training, but also to spontaneously show this ability in novel situations with such sophistication that it seems unlikely that they used processes that purely rely on estimation (Pepperberg, 2012; Cantlon, 2012).

The ability to discriminate quantities seems to be widespread in the animal kingdom and there are common constrains regarding the patterns of performance (Cantlon, 2012). In many species it has been shown that the accuracy of performance increases with the difference between the two quantities, which holds true for both mammals (*apes*: Beran, 2001; *dogs*: Ward & Smuts, 2007; *rhesus monkeys*: Jordan & Brannon, 2006) and birds (*pigeons*: Emmerton & Renner, 2006; *African Grey Parrots*: Aïn et al., 2008).

Weber`s Law, also known as *numerical ratio effect*, is an important behavioural signature of *numerousness judgements* or *numerosity discrimination* (Gallistel & Gelman, 2005). It describes a dependency on the ratio between values that are being compared in order to psychologically discriminate them (Pepperberg, 2012). This applies only to approximate quantity discriminations as opposed to precise ones, and also implies that there is a certain error in psychological representation that increases with the quantity (Cantlon & Brannon 2010). In an experiment where the task for the subjects was to choose the smaller amount of two different quantities, humans and monkeys performed similarly especially when regarding that for both the response time increased and accuracy decreased, as the ratio between two amounts approached 1:1 (Cantlon & Brannon, 2010). It has been observed that quantification below amounts of four items is generally done more accurately and uses less reaction time compared to larger sets that represent the same ratio between the tested amounts. Thus it has been hypothesized that *Weber`s Law* does not suffice as explanation of quantification of fewer items, thereby indicating the existence of a unique cognitive mechanism that is used in *subitizing* (Piazza et al., 2011). The results of the study conducted for this purpose provided evidence that discrimination difficulty predicted by *Weber`s Law* does not account for discriminating quantities in the low range, making a unique system – or at least a " … special treatment within the estimation system … " – in a higher range plausible (Revkin et al., 2008).

Introduction

It has been argued that adaptive pressure fosters the development of quantity discrimination in many different contexts across diverse species (Beran et al., 2015). For example, both predator avoidance strategies and tactics for locating food have been suggested to benefit from the ability to discriminate quantities (Beauchamp, 2003; Farnsworth & Smolinski, 2006; Anobile et al., 2015). An example for the former could be enhanced survival due to knowing when flight is necessary because of being outnumbered by predators. Regarding the latter, animals having this ability would benefit in evaluating which food source is richer and delivering more energy due to yielding more portions compared to other ones. This means an advantage over animals that are not as skilled in quantity discrimination. Optimal foraging theory states that when faced with alternate options while foraging, animals will choose the one with the highest return rate of food (Pulliam, 1974). In order to demonstrate how the ability to discriminate quantities can provide a benefit in this regard, a study has been conducted in which northern mocking birds had to remove different amounts of wooden sticks to get a reward (Farnsworth & Smolinski, 2006). They chose the option were the least amount of obstacles were given significantly above chance level, thereby solving the task and obtaining food resources at a high rate on account of their ability to discriminate quantities (Farnsworth & Smolinski, 2006).

1.3. Empirical studies of quantity discrimination

There has been a diversity of experimental setups in order to test quantity discrimination abilities in animals. An interesting discourse in accounting for quantity discrimination abilities lies in the way of testing the different sensory modalities that can be involved in the process. For example, instead of discriminating visual aspects, sound pulses could make up the stimuli. In this case they can only be perceived sequentially and the delay in time is a factor that has to be controlled for while conducting the experiment and recognized when analysing the data. In some cases mental representations of quantities are part of the experimental design, by having the subjects choose from memory of either acoustic (Abramson et al., 2013) or visual information (Range et al., 2012; Beran et al., 2013; Ujfalussy et al., 2013). Honeybees had been trained to locate a food source placed after a certain number of coloured tents used as landmarks. By varying the traveling distance necessary to get to the feeding spot, it was shown that the bees rely on the number of landmarks rather than distance, demonstrating their numerical competence and sequential memory capacity (Chittka & Geiger, 1995).

In the majority of studies researching this topic, however, two-way choice situations are created, meaning two stimuli are presented that the subjects have to choose from

Introduction

(Watanabe, 1998; Petrazzini, 2014). Another type of discrimination task entails the subjects having to choose the correct stimulus out of multiple possibilities to match up with a given stimulus presenting a specific quantity, such as in a delayed matching-to-sample (DMTS) task (Boysen & Berntson, 1989; Vonk, 2013; Pahl et al., 2013). In visual discrimination tests the stimuli to discriminate range from real-life objects to representations on a touchscreen (Vonk, 2013). The elements comprising the stimuli can be presented in a static or moving state (Vonk & Beran, 2012), in one study even being live conspecifics used as display of quantities (Agrillo et al., 2009). Depending on the design, different aspects need to be controlled, for example when using continuous flashes of light as indicator of quantity the temporal dimension has to be controlled for. In order to make sure that any subject in a quantity discrimination experiment succeeds in the task due to relying on its skill for quantity discrimination and not on skills utilizing other parameters, such as the timespan that the stimuli can be heard during an experiment based on audio playback. When utilizing graphical stimuli, the summed surface area of the elements has to be controlled in order to avoid providing information about the quantity the elements represent (Watanabe, 1998). Oftentimes the reinforcement is delayed – as it is in this study – but sometimes the reward is the stimulus itself and can be obtained by the subjects at the end of a trial (Ujfalussy et al., 2013; Rahman et al., 2014; Ward & Smuts, 2007).

The ability for quantity discrimination has been found in a wide range of species. Not only has it been shown to exist in mammals (dogs: Ward & Smuts, 2007; wolves: Range et al., 2012; horses: Petrazzini, 2014; sea lions, dolphins, beluga whales: Abramson et al., 2011; Abramson et al., 2013; rhesus macaques: Cantlon & Brannon, 2007; chimpanzees: Beran, 2001; Beran et al., 2013) and various birds (jackdaws: Ujfalussy et al., 2013; chickens: Rugani et al., 2007; pigeons: Watanabe, 1998; Emmerton & Renner, 2006; Hirai, S., Jitsumori, 2009; African grey parrots: Aïn et al., 2008; Pepperberg, 2006) but also in fish (mosquitofish: Agrillo et al., 2009), amphibians (salamanders: Uller et al., 2003; Krusche et al., 2010) and insects (honeybees, bumblebees: Pahl et al., 2013). While all of these species showed quantity discrimination capabilities, the quantities they were able to distinguish successfully varied greatly. The methodologies used to show this differ, and the range of quantities – meaning the amounts between the lower and upper limit – as well as the increments – meaning the difference between the values of quantity to discriminate – vary. All species tested needed to use their ability to choose either the higher or lower amount of a given set of stimuli above chance level.

Two groups of mammals that have been repeatedly tested are canids and primates. An experiment with dogs (*Canis lupus familiaris*) showed that they could discriminate the

Introduction

quantity combinations of 1 vs. 4, 1 vs. 3, 2 vs. 5, 2 vs. 5 and 3 vs. 5 significantly above chance level but could not do so when the combination pairs differed by only one increment, as in the combinations 1 vs. 2, 2 vs. 3 and 4 vs. 5 (Ward & Smuts, 2007). Wolves (*Canis lupus*) on the other hand were able to discriminate quantity stimuli sets of 1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 4, 2 vs. 3 and 3 vs. 4 above chance level, even by relying only on mental representation of the stimuli objects as they were not presented all at once. Instead they have been shown sequentially one after another with their summed amounts having been visually obstructed altogether (Range et al., 2012). Regarding primates, an orang-utan (*Pongo abelii*) demonstrated the ability for quantity discrimination in a matching-to-sample task, correctly choosing the right stimuli in the combination sets of 1 vs. 2, 3, 4 as well as in 4 vs. 5, 6, 7 and 7 vs. 8, 9 and 10 (Vonk, 2013). Two western lowland gorillas (*Gorilla gorilla gorilla*) were tested in two-way choice tasks using a touchscreen with either static or moving dots and even moving sub-sets of dots as stimuli. They were able to discriminate all the combinations possible between the amounts 1 to 10 with at least one increment difference for the static and moving stimuli conditions above chance level (Vonk et al., 2014). In an experiment with four chimpanzees (*Pan troglodytes*) the aim was to test if they could choose the larger quantity of two possibilities in a set-up were the stimuli were individually hidden beneath opaque cups. These cups were ordered in two rows, with the subjects having to choose the row that has more cups containing a food item. They chose significantly above chance level in all the presented combinations of 1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 3, 2 vs. 4 and 3 vs. 5 in the first phase and combinations 1 vs. 3, 2 vs. 3, 2 vs. 4, 3 vs. 4 and 3 vs. 5 in the second phase (Beran et al., 2013). In some studies even the ability of non-human primates to use simple arithmetic has been tested, showing that they indeed were able to add small quantities by correctly choosing their sum (rhesus macaques: Boysen & Berntson, 1989; Cantlon & Brannon, 2007; Cantlon, 2012).

The very first experiments in order to identify the ability for quantity discrimination in birds go back almost a century ago to the work of Otto Köhler, who investigated this ability in pigeons, parakeets, jackdaws, magpies, ravens and grey parrots amongst other animals (Köhler, 1941; 1943; 1951). Many experiments testing birds in this regard have been conducted since. The New Zealand robin (*Petroica australis*), a food-hoarding songbird, showed the ability to discriminate between the quantities 1 vs. 2, 2 vs. 3, 3 vs. 4 and 4 vs. 6 above chance level, while not being able to discriminate combinations 4 vs. 5, 4 vs. 8, 6 vs. 8, and 8 vs. 10, when controlled for time and volume confounds (Hunt et al., 2008). When jungle crows (*Corvus macrorhynchos*) were presented a choice test between two sets of discrete homogeneous quantities, they were able to discriminate the sets 3 vs. 5, 4 vs. 5, 5 vs. 7 and 5 vs. 8, but not that of 5 vs. 6 quantities significantly above chance level (Bogale et al.,

Introduction

2011). Jackdaws (*Corvus monedula*) showed their ability for quantity discrimination in a two-way choice task using mental representation as the stimuli had been placed sequentially into two non-transparent containers. They chose correctly significantly above chance level for the combinations 1 vs. 2, 1 vs. 3, 2 vs. 3, 1 vs. 4, 2 vs. 4, 1 vs. 5 and 2 vs. 5, but failed to do so for the combinations 3 vs. 4, 3 vs. 5 and 4 vs. 5 (Ujfalussy et al., 2013). In an experiment involving three African grey parrots (*Psittacus erithacus*), one of them chose the larger of two amounts for the combination pairs 1 vs. 2, 1 vs. 3, 1 vs. 4, 1 vs. 5, 2 vs. 3, 2 vs. 4, 2 vs. 5, 3 vs. 5 significantly above chance level, whereas the other one failed to do so for the combination pairs 3 vs. 4 and 4 vs. 5 and the third one for the combination 4 vs. 5 (Aïn et al., 2008). Pepperberg famously tested a specific African grey parrot named Alex in diverse experimental setups involving quantity discrimination (Pepperberg, 1994; 1999; 2006). One such study consists of two consecutive experiments, in the first of which Alex was shown two sets of items. Each set consisted of one to six items. The sets differed in colour, but all items of one set matched in colour and object type. The parrot was then asked "What colour X?" – whereby X the amount in question is – and had to vocalize the correct colour word. The results were significantly above chance level, with 32 correct answers out of 35 trials. In the following experiment Alex was shown three sets of items consisting of different amounts and either different colours or different object types. In this experiment his objective was to vocalize either the colour or the material respectively, that was fitting the asked number tag. In this experiment Alex was correct significantly above chance level as well, with 66 correct answers out of 74 (Pepperberg & Gordon, 2005).

Although in non-human animals the ability to respond to varying dimensions of quantity has been studied in much detail and analogies between basic cognitive mechanisms across a myriad of animal species have been established, the exact underlying processes they utilize for this purpose are still not entirely understood (Davis & Pérusse, 1988; Cantlon, 2012). The hypothesis that quantity discrimination is easier when the elements of the stimuli to discriminate are positioned based on the structure of a symmetrical grid has not been tested in animal cognition experiments. In human psychology, a study was conducted that found that when arrays of dots were ordered according to visual pattern opposed to being randomly distributed, the reaction time dropped and range of quantities yielding correct choices was improved (Krajcsi et al., 2013). Another study tested the effect of pattern based quantity stimuli enumeration on children (Jansen et al., 2014). In this experiment school children were tested on their ability to enumerate the exact amount of dots displayed either in a random placement, a dice pattern or a line configuration for the quantities 1 to 6. The results demonstrate that the main effect of configuration of stimuli elements was significant and that quantities 5 and 6 on the dice pattern were significantly easier to enumerate for the children

compared to the other conditions. Pepperberg points out, referring to other studies, that children above the age of 3 and a half years typically understand number labels fully and are able to count, but the African grey parrot she tested exceeded their ability below that age. In this paper, Pepperberg also holds that parrots generally might have better perceptual mechanisms compared to young children (Pepperberg & Gordon, 2005). To fill the scientific gap and contribute knowledge about the effect of placing the elements of stimuli according to a grid pattern, this will be amongst the experiments that the kea participate in during this study. The research questions and hypotheses will be presented in detail during the following subchapter.

1.4. Research questions and hypotheses

Research question 1a: When kea (*Nestor notabilis*) have to discern between two stimuli displaying different quantities, does their pattern of performance change when the relative difference between two amounts compared to their overall size is increased?
Research question 1b: Is the performance of Kea in choosing the higher quantity of a stimulus pair determined by the absolute or relative quantity difference between the stimuli?
To test for this, the absolute difference in quantity has been kept constant (the increments ranging from one to four), whereas the relative difference was changing dynamically by increasing the amounts presented. To illustrate this utilizing an example from the quantities used, when comparing the quantity 7 with 10, the absolute difference is 3 whereas the relative difference is 0.7. When comparing the quantities 19 and 22, the absolute difference is still 3 but the relative difference is at ~0.86.

Ha(1a): The relative difference in number of objects between two sets affects the kea"s ability to discriminate between them.
H0(1a): The relative difference in the number of objects between two sets does not affect the kea"s ability to discriminate between them.
Ha(1b): The performance of Kea in choosing the higher quantity of a stimulus pair is determined either by the relative or the absolute quantity difference between the stimuli.
H0(1b): The performance of Kea in choosing the higher quantity of a stimulus pair is neither determined by the relative, nor the absolute quantity difference between the stimuli.

Prediction for research question 1a and 1b: Kea will perform better at choosing the higher of two sets of quantities presented, as the difference between their amounts is increased relative to the size of their amounts and their performance is determined by the relative quantity difference between the stimuli.

Introduction

Research question 2a: Is the pattern of performance shown by kea in choosing the higher quantity depicted by one of two stimuli influenced when the stimuli presented are two-dimensional pictures compared to being real three-dimensional objects?

Research question 2b: Is the pattern of performance shown by kea in choosing the higher quantity of a stimulus pair determined by the absolute or relative quantity difference between the stimuli?

Ha(2a): In choosing the higher quantity depicted by one of two stimuli, the pattern of performance shown by kea is similar when stimuli are real three-dimensional objects compared to if they are virtual two-dimensional objects displayed on a touchscreen.

H0(2a): The pattern of performance by kea in choosing the higher quantity depicted by one of two stimuli is different when the stimuli are presented as real objects compared to if they are virtual two-dimensional objects displayed on a touchscreen.

Ha(2b): The performance of kea in choosing the higher quantity of a stimulus pair is determined either by the relative or the absolute quantity difference between the stimuli if they are virtual two-dimensional objects displayed on a touchscreen.

H0(2b): The performance of kea in choosing the higher quantity of a stimulus pair is neither determined by the relative, nor the absolute quantity difference between the stimuli if they are virtual two-dimensional objects displayed on a touchscreen.

Prediction for research question 2a and 2b: The pattern of performance by kea in choosing the higher quantity depicted by one of two stimuli is increased when the stimuli are presented as real objects compared to if they are two-dimensional virtual objects displayed on a touchscreen. Their performance is determined by the relative difference between quantities.

Research question 3: Does a grid structure arrangement of graphic elements of two stimuli displayed on a touchscreen affect the kea‟s ability to discriminate quantities?

Ha(3): A grid structure arrangement of stimuli elements affects the kea‟s ability to discriminate quantities.

H0(3): A grid structure arrangement of stimuli elements does not affects the kea‟s ability to discriminate quantities.

Prediction for research question 3: In discerning between two sets displaying different quantities on the touchscreen, kea perform better at choosing the larger quantity when the organization of graphic elements composing the presented stimuli is based on the structure of a grid in contrast to them being placed randomly.

2. Materials and Methods

2.1. Subjects: Kea (*Nestor notabilis*)

The kea (*Nestor notabilis*) is a species of parrot (*Psittaciformes*) endemic to the alpine regions of New Zealand, first having been described in the journal of English and foreign literature, science and fine arts named "*Athenaeum*" in April 1856 and three months later by ornithologist John Gould in a paper published by "*Proceedings*" (McAllan, 2004). The term for the genus "*nestor*" is a reference to Nestor, the king of Pylos who fought in the Trojan Wars and having been praised in Homer`s Iliad for being "great-souled" and "clear-voiced" and the species name literally translates to "remarkable", reflecting the birds unique attributes (Fraser & Gray, 2013). The common name "kea" originates from the Māori language and is onomatopoetic, referring to the screech-sound produced by the birds (Toft & Wright, 2015). The kea is one of the few parrot species inhabiting alpine regions. It is native to the mountains of South Island, New Zealand and can be found from Fiordland north to Nelson and Marlborough Provinces in habitats characterized by montane woodlands, highland valleys, subalpine scrublands and alpine grasslands, an environment most frequently found in ranges of altitude between 600 and 3000 metres (Diamond & Bond, 1999; Forshaw, 2010).

Their appearance is characterized by their olive-green colour and orange-red coverts from back to rump and under-wings. The undersides of their flight feathers are of orange-yellow colour (Diamond & Bond, 1999; Forshaw, 2010)*.* They have been described as being noisy, conspicuous, highly inquisitive and as unique in their appearance (Forshaw, 2010). Kea are often being observed near ski lodges, disassembling parked cars, tents and similar constructions and thus generally known for their boldness (Brejaart, 1988).

Kea possess high levels of intelligence and flexibility, are curious, persistent and display the most extensive and elaborate play behaviour of any known bird species. The authors Diamond and Bond view the typically highly neophilic and inquisitive character of kea as what they call "flexibility as a lifestyle", and claim this to be a necessary evolutionary development posed by influences of their adamant environment (Diamond & Bond, 1999). Kea engage in complex social behaviour, also having implications for the creation of orderings in feeding priority. Like in other parrots, the young are fed by their parents utilizing regurgitation until they progressively learn to be self-sufficient (Diamond & Bond, 1999).

Contrary to most other parrot species, kea are omnivorous, feeding on plants, eggs, insects, chicks, mice, reportedly even on sheep and carrion of red deer and possum (Brejaart, 1988; Schwing, 2010). This also led to them being hunted for fear of attacks on livestock, following the introduction of sheep to the high country of the south island of New Zealand (Meyers, 1924; Marriner, [1908] 2010). The conservation status of kea can be described as vulnerable, which means that they are at high risk of endangerment in the wild (Forshaw, 2010). The historical conflict with humans, including the threat of poaching and continuing habitat destruction, caused their conservation status (Diamond & Bond, 1999).

2.1.1. Housing conditions of the Haidlhof Research Station

For the experiments of this master`s thesis captive kea (*Nestor notabilis*) housed at the Haidlhof Research Station in the vicinity of Bad Vöslau, Austria, have been tested between March 16th, 2015 and June 25th, 2015. The experiments involved a maximum of 13 kea, group-housed in the aviary of 520 m^2, where all of the experiments took place in a separable compartment. Food consisting of fruits and vegetables and diverse seeds was delivered thrice daily and additionally a source of protein was fed once per day. Water was available *ad libitum,* also during experimentation.

2.1.2. Ethical considerations

The experiments for the master`s thesis have been discussed and approved by the institutional ethics and animal welfare committee in accordance with GSP guidelines and national legislation of the University of Veterinary Medicine, Vienna (ETK-03/08/2015), and were not classified as animal-experiments under the Austrian Animal Experiments Act (TVG 2012). The subjects that participated in the experiments are housed in accordance with the Austrian Federal Act on the Protection of Animals (Animal Protection ActdTSchG, BGBl. I Nr.118/2004). Furthermore, as the present study was strictly non-invasive and based on behavioural observations only, none of the experiments were classified as animal-experiments under the Austrian Animal Experiments Act (x 2, Federal Law Gazette No. 501/1989), requiring further permission.

During testing we have done our best not to stress the birds and to acknowledge when they were not willing to participate in the experiments. It was necessary to factor their rank into the order of participation, because when a bird of high rank wanted to participate first, it would tend to stress the lower ranking birds by chasing them away when they were trying to be tested before. When the weather was getting very hot in summer, the kea would not participate in the touchscreen task of Experiment I stage II and Experiment II, as they were

not motivated enough by the food reward. For the second Experiment one bird (Papu) had to be excluded, as she was not willing to participate, and another bird (Elvira) due to being separated from the other birds because of medical issues. Every day before testing, the aviary was cleaned by me and the caretakers. Only high quality organic peanuts were used as reward. To paint the wooden balls utilized as part of the stimuli, only food colouring was used. Testing took place between 10 and 12 a.m. and between 2 and 4 p.m., so as not to disturb the feeding time, while providing an additional hour of rest and digestion after feeding.

2.1.3. Subjects of Experiment I stage I

For the first stage of Experiment I, the sample included Anu, Elvira, Kermit, Roku and Pick. Anu, Kermit and Pick are male and have been raised by hand. Roku is male and has been parent raised and Elvira is female and also has been parent raised. All birds had prior experience with laboratory experiments in general and had been trained in two-way choice tasks in particular. The five subjects participating in this experiment also all had prior experience with quantity discrimination tasks, as they performed in a very similar experiment conducted for a pilot study, with the difference that only lower quantity combinations had been tested. Testing for stage I of Experiment I took place between 16th of March 2015 and 11th of May 2015.

2.1.4. Subjects of Experiment I stage II

For the second stage of the first experiment, 13 kea were used as subjects, namely Papu, Paul, Lilly, Willy, Sunny, Coco, Frowin, John, and the five kea from the first part of the experiment, namely Anu, Elvira, Kermit, Roku and Pick. Coco, Willy, Lilly and Sunny are female and were handraised. Paul, John, Frowin are all male and have been parent raised, only Papu is female and was handraised. None except for the five subjects participating in stage I of Experiment I had prior experience with quantity discrimination tasks. Testing for stage II of Experiment I took place between 1th of June 2015 and 6th of June 2015.

2.1.5. Subjects of Experiment II

The subjects for Experiment II included 11 kea, namely Paul, Lilly, Willy, Sunny, Coco, Frowin, John, Anu, Kermit, Roku and Pick. They all participated in Experiment I part II already, but from the 13 originally ones tested, Papu had to be excluded because she did not want to participate in the experiment and Elvira had to be excluded, because she was having medical problems. All eleven kea had prior experience with quantity discrimination tasks as

they were participating in stage I and II of Experiment I. Testing for Experiment II took place between 15th of June 2015 and 24th of July 2015.

2.2. Stimuli

2.2.1. Stimuli of Experiment I, stage I

For stage I of the Experiment I, the stimuli consisted of wooden panels (app. 30 cm x 30 cm), with green coloured wooden balls (diameter Ø 2.5 cm, using food colour as dye) attached to them (Fig. 1). We chose green as it is the least distracting colour for kea (Weser & Ross, 2013). Mounts were constructed so that these panels could be vertically plugged in wooden sockets of approximately 2 cm height. The stimuli were placed at an equal distance of about 3 metres away from the entrance to the testing compartment and approximately 1 metre apart from each other. The mounted plates were angled in such a way, that the kea had a full frontal view of the stimuli while the reward (pieces of organic food peanuts) stayed out of the bird`s sight as it was placed on the wooden mount behind the plate. In similar studies the reward and stimuli are unified, but in the case of this study it would not have been feasible, as the quantities tested were too high to ensure motivation of the kea being kept up for an entire session.
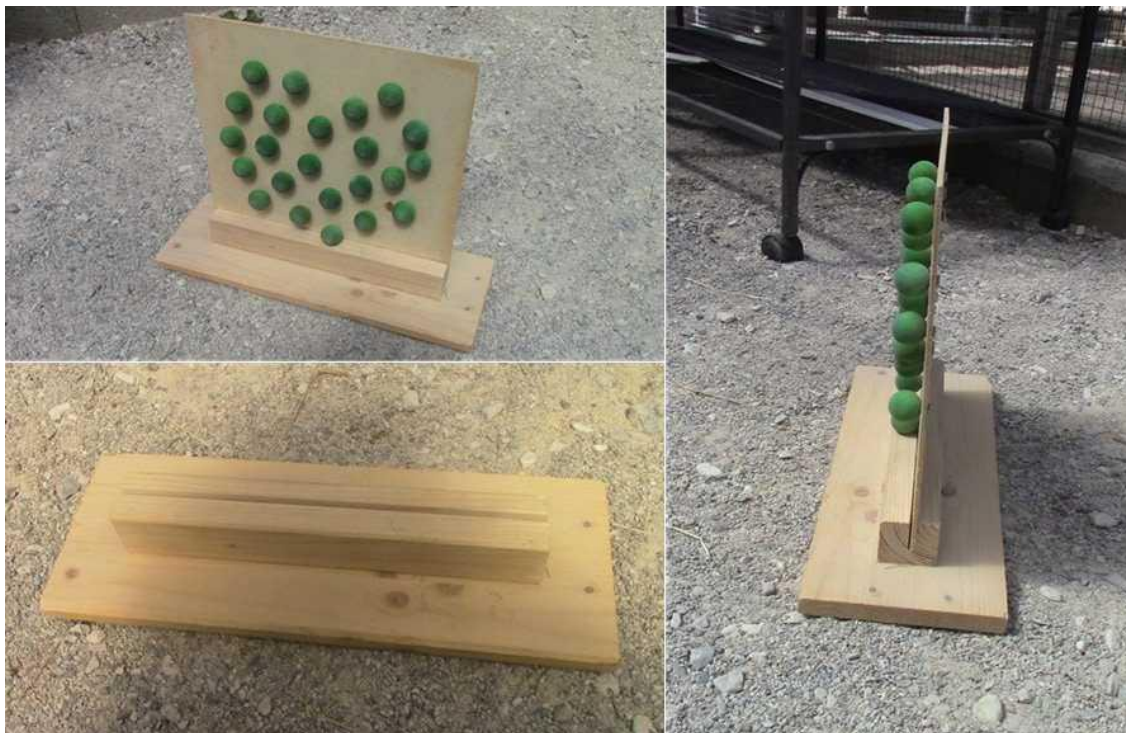

Fig 1. Upper Left: Frontal view of stimulus with elements approx. 1-2cm apart. Right: Sideways view of stimulus with peanut reward behind the plate. Lower Left: Mounting base of stimuli.

Materials and Methods

For every quantity displayed two stimuli were constructed in order to control for the Kea being able to solve the task by comparing the overall area that the elements of the plates take up rather than the quantity of them. On one version of stimuli, the elements were mounted closer to each other, at approximately 1-2 cm, and on the other they were mounted further apart from each other, at about 3-4 cm. For both versions the elements were mounted in a random orientation to one another on the plates. As the plates were square-cut they could be placed on either one of their four sides on the mount, thus additionally allowing for 4 different orientations per stimulus.
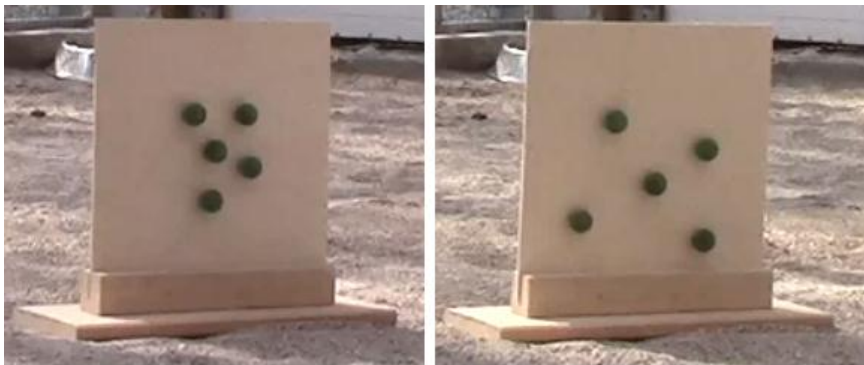

Fig 2. Example of the two versions of a stimulus depicting the quantity of 5, but with closer spaced elements (left) and elements spaced further apart (right).

2.2.2. Stimuli of Experiment I, stage II

During the second stage of the first experiment, as well as the second experiment, a touch-screen as part of the "*Vienna comparative cognition technology*" (VCCT) was used (Fig. 11, 12). It is an operant conditioning system that has been applied in studies on diverse species, such as dogs, wolves, pigeons, corvids, kea and humans (Steurer et al. 2012). A flatscreen display is used to present the stimuli with a display area of 304 mm × 228 mm (381-mm diagonal) and a resolution of 1,024 × 768 pixels, each pixel thus having a size of 0.297 mm × 0.297 mm. Responses are detected by an 15 inch infrared touch-frame using "Carroll Touch Technology", with a safety-glass plate placed behind the infrared array. Being placed at a distance of 7 mm to the beams, the surface of detection is shifted towards the base instead of the apex of the bird`s beak, thereby allowing for an enhanced validity in the recognition of pecks (Steurer et al. 2012).

The stimuli used were bitmaps created using Flash Professional CS6 (version 12.0.2, Adobe Systems Incorporated, San Jose, California, USA), that show a fixed light brown (web colour: #DCC761, most closely resembling web safe colour: #CCCC66, equals RGB: 220, 199, 97) coloured square with different amounts of green (web colour: #3EFE05, most closely

resembling web safe colour: #33FF00, equals RGB: 62, 254, 5) coloured squares placed inside its boundaries. In order to control that the kea were choosing on the basis of the quantity of elements of the stimuli and not the relative or absolute surface area they make up, the squares were constructed using an identically sized surface area (60*60 px) as basis for all stimuli used in stage II of Experiment I (see Appendix I.I.I and I.I.II). When the same surface area is transformed into two sets of differing quantities, more elements would mean that they each are comparably smaller and less elements that they would be comparably larger. In order to control for this, the elements comprising a single stimulus differ in size from each other, so that the kea cannot rely on the smaller squares representing the larger quantity.

For this stage of the experiments, the squares were distributed randomly inside the borders of the background. Also, the difference in size between squares was controlled so that the largest square was less than double the size of the smallest square in any pair of stimuli presented together. Every stimulus has been constructed in two versions that were presented semi-randomized, so that the effect of association was decreased.
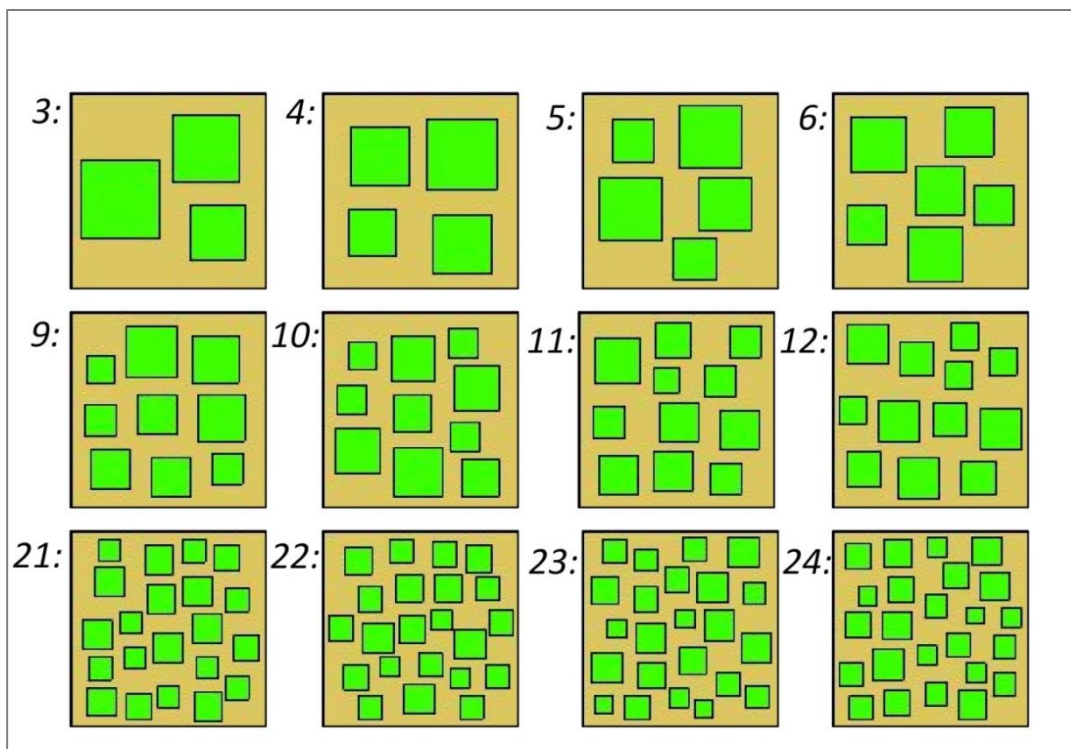

Fig 3. Example of "virtual" stimuli used in Experiment I, stage II.
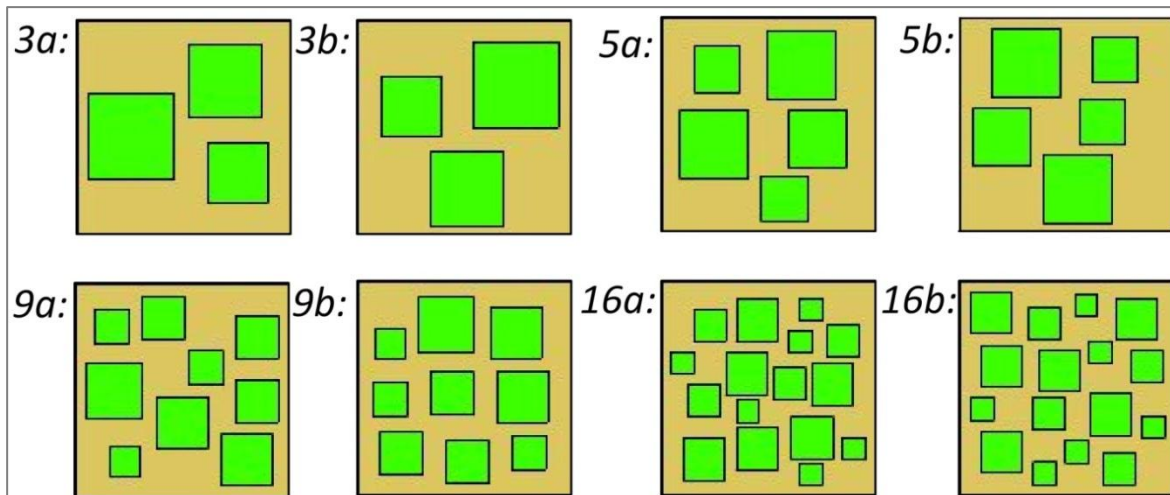
Fig 4. Example of two stimuli versions (a / b) depicting the quantities of 3, 5, 9 and 12.

2.2.3. Stimuli of Experiment II

The stimuli used in Experiment II were also virtual presentations on the VCCT (bitmaps created using Adobe Flash CS6), but the elements of them were placed on fixed symmetrical positions, locked on an invisible grid (for all stimuli used in Experiment II, see Appendix I.II.I and I.II.II). The stimuli show a fixed light brown coloured square with different amounts of green coloured squares placed inside its boundaries, using the same exact colour as in Experiment I part II. Every stimulus was constructed in two versions that were presented semi-randomized in order to decrease the effect of association.

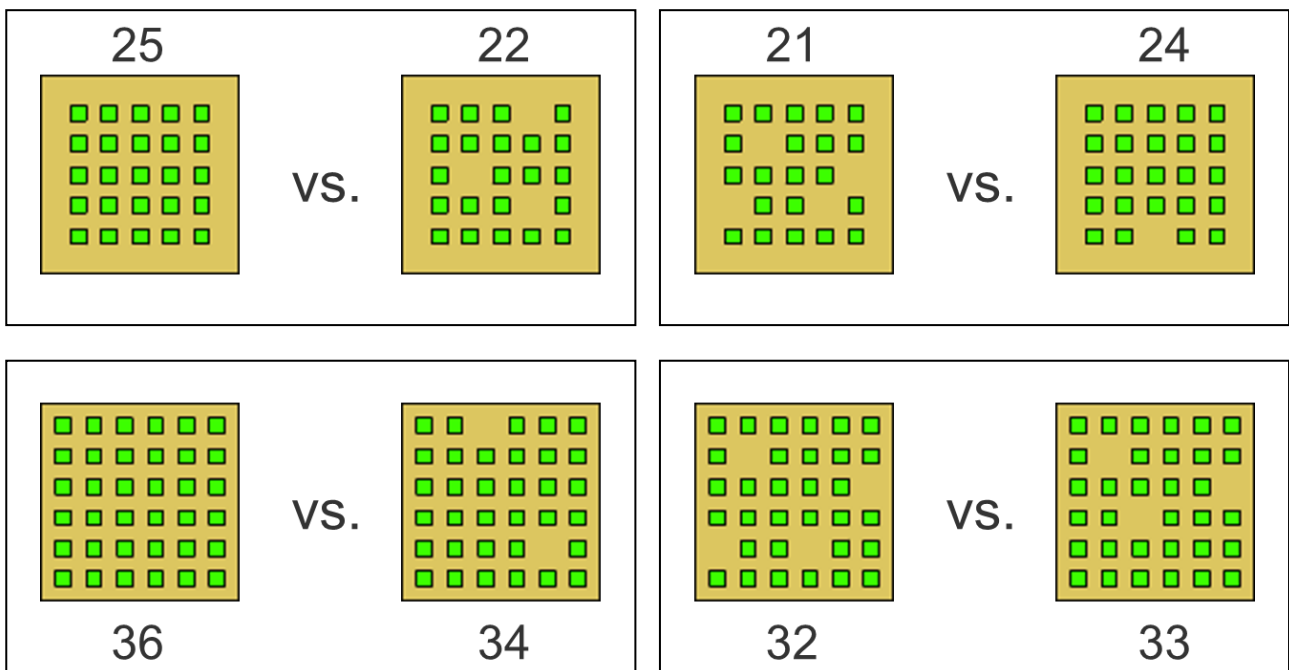**Fig 5. Examples of four matched pairs of virtual stimuli in Exp II.**
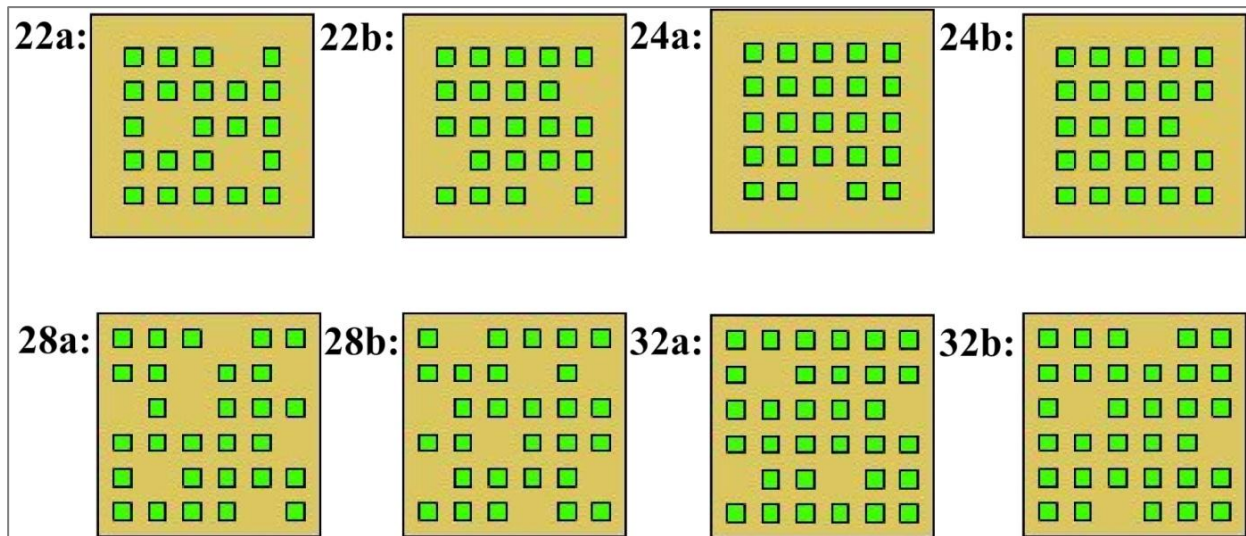
Fig 6. Example of two versions (a / b) of stimuli depicting the quantities 22, 24, 28, 32.

2.3. Procedure

2.3.1. First stage of Experiment I

The research aim of the first stage of Experiment I was to show if kea rely on the absolute or relative difference in their ability to discriminate between two unequal amounts in a definite set of quantities. To test for this, the absolute difference in quantity was kept constant, the increments ranging from one to four, whereas the relative difference was changing dynamically by increasing the amounts presented (stepwise increments from 5 vs. 6, 7, 8, 9 up to 20 vs. 21, 22, 23, 24). This resulted in 64 combinations, with each of the five birds having 10 trials per combination, summing up to 50 trials per combination, 640 trials per bird and 3200 trials in total for all birds and combinations. The kea were tested individually 32 trials per session with a maximum of 2 sessions per day, once antemeridian and once in the afternoon. The "lower quantity combinations", namely the quantities from 5 vs. 6 to 12 vs. 16 have been tested before increasing the quantities to the "higher quantity combinations" of 13 vs. 14 up to 20 vs. 24, and during both sets of possible pairs a fully randomised order of combinations was provided.

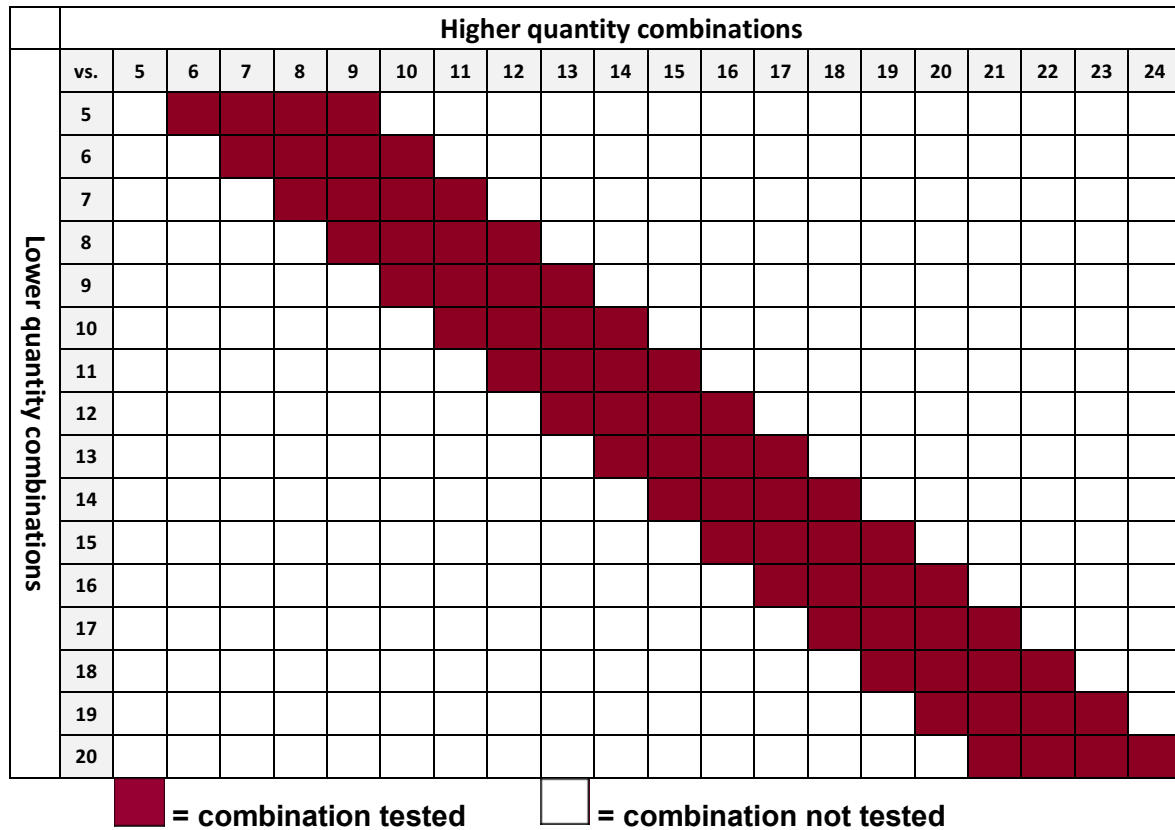|		Higher quantity combinations																		
	vs.	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Lower quantity combinations	5		■	■	■	■															
	6			■	■	■	■														
	7				■	■	■	■													
	8					■	■	■	■												
	9						■	■	■	■											
	10							■	■	■	■										
	11								■	■	■	■									
	12									■	■	■	■								
	13										■	■	■	■							
	14											■	■	■	■						
	15												■	■	■	■					
	16													■	■	■	■				
	17														■	■	■	■			
	18															■	■	■	■		
	19																■	■	■	■	
	20																	■	■	■	■

■ = combination tested   □ = combination not tested

Fig 7. Illustration of quantity combinations used in stage I of Experiment I.
Red squares demarcate the quantities used as a pair in a trial. The upper row depicts the higher quantity of the pair, increasing from left to right by one increment per square. The column to the left depicts the lower quantity of the pair, increasing downwards by one increment per square.

For the experiment a semi-randomised side of reward was used, so that each side was at most the correct side three trials in a row before changing the reward side. This was done to reduce the possibility of reinforcing a side-bias. The criterion of exclusion due to side-bias was set at 85%. For training only combinations in the low range have been used that were not presented during testing itself, namely the combinations 1 vs. 6, 7, 8; 2 vs. 7, 8 and 3 vs. 8, with a criterion of 17 correct choices out of 20 trials in two consecutive sessions. For training the kea were tested individually 32 trials per session with a maximum of 2 sessions per day, once antemeridian and once in the afternoon, for 3 days per week until they reached the criterion.

In the first stage of Experiment I, the kea had been individually led from the main aviary into the experimental compartment ("*porticular res*") of the kea aviary. There they were separated from the experimental area by a meshed sliding door, which was closed. The experimental compartment and the experimental area both were visually separated from the remaining

aviary by a set of white plastic sliding plates. All the stimuli lay in the corner of the experimental area face down so that the kea would not be distracted by them. Before each trial, the experimenter placed the stimuli according to the test sheet. He then placed the peanut behind the stimuli displaying the higher quantity.

Even though the keen sense of smell in kea is scientifically proven (Gsell et al., 2012) we did not control for olfactory cues as the factors of distance, wind and the wooden plate of the stimuli seemed to be sufficient natural obstacles as to interfere with the kea relying on their olfactory sense to determine their choice. In order that the kea could not see where the reward was put, the peanut was palmed by the experimenter and he simulated the movement with the opposite hand, displaying the same gesture behind both stimuli. Subsequently, the experimenter opened the sliding door, letting the bird into the experimental area, and walked straight behind the stimuli so that the kea was not influenced in its choice. The bird then could choose between the two different stimuli. It achieved this by walking behind the wooden plates comprising the stimuli to retrieve the reward of a quarter piece of peanut behind the stimuli displaying the higher quantity. The stimuli were angled sideways (see Fig. 8) in such a way that they kea was not able to walk through between them looking for the reward. The bird had to look behind a specific plate so that the choice was always obvious to the experimenter. After each attempt, the bird had to return outside the experimental area.

Fig 8. Setup of first stage of Experiment I in "porticular res". Stimuli are placed facing downwards in-between trials to avoid being seen by the subjects; Mounting bases are positioned as in the experimental procedure.

Fig 9. Positioning of stimuli and kea during first stage of Experiment I. Before the
door is opened to enable the kea to choose one of the two stimuli by approaching it.

Fig 10. Sight barrier of "porticular res".

2.3.2. Second stage of Experiment I

The second stage of the first Experiment entailed transmitting the task onto the touchscreen (VCCT). The kea were presented with two stimuli displaying different quantities and had to choose the one displaying the higher amount by touching the screen (see Fig. 12). They retrieved a piece of peanut after each correct choice automatically from the dispenser of the touchscreen unit.

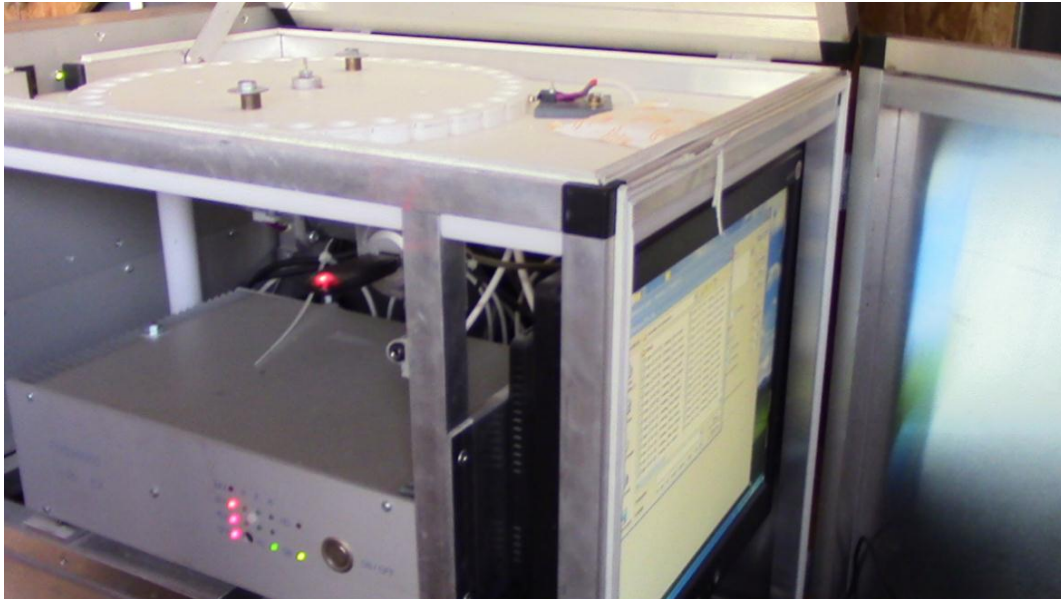**Fig 11. Sideways view of VCCT.**
  Reward dispensing wheel at the top; Monitor for the experimenter;
  CPU at the left bottom.

Fig 12. Upper-left view of VCCT touchscreen monitor.
  Displaying two stimuli of the second stage of Experiment I.

The aim of this stage of Experiment I was to test if absolute or relative difference between the two stimuli presented on the touchscreen is correlated with success of the kea. For training, only combinations in the low range were used that were not presented during testing itself, with a criterion of 17 correct choices in two consecutive sessions consisting of 20 trials per session. The tested quantity combinations utilized in the second stage of Experiment I were in the range from 1 vs. 2 up to 20 vs. 24, including the combinations provided in part I of Experiment I. This led to 20 combinations, with each of the 13 birds having 10 trials per combination, summing up to 130 trials per combination, 200 trials per bird and 2600 trials in total for all birds and combinations. The kea were tested individually 32 trials per session with a maximum of 2 sessions per day, once antemeridian and once in the afternoon. A fully randomised order of combinations was provided, and due to the architecture of the software, the rewarded side was also fully randomized. Both of the two stimuli versions were utilized for an equal part during testing. Regarding a difficulty for the birds to generalize from real life objects to the touch screen, this has been tested with the result of them successfully demonstrating both picture-to-object and object-to-picture transfer, performing at a comparable level when pictures were displayed on a touchscreen or as printed photographs (Wein et al., 2015).

Higher quantity combinations (columns labelled 1–24) vs. **Lower quantity combinations** (rows labelled 1–20). Legend: ■ = combination tested; □ = combination not tested; ▨ = tested in Exp. I stage I.

Vs.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1		■	▨	▨	▨																			
2			▨	■	▨	▨																		
3				▨	▨	■	▨																	
4					▨	▨	▨	■																	
5					■	▨																			
6						▨	■	▨	▨																
7								▨	▨	■	▨													
8									▨	▨	▨	■												
9										■	▨	▨	▨											
10											▨	■	▨	▨										
11												▨	▨	■	▨									
12													▨	▨	▨	■								
13														■	▨	▨	▨							
14															▨	■	▨	▨						
15																▨	▨	■	▨					
16																				■				
17																		■	▨	▨	▨			
18																				■	▨	▨		
19																					▨	■	▨	
20																								■

■ = combination tested   □ = combination not tested   ▨ = tested in Exp. I stage I

Fig 13. Quantity discrimination pairs of stimuli used in Experiment I stage II.

2.3.3. Experiment II

For Experiment II the organization of the graphic elements of the digital, bi-dimensional stimuli depicting the quantities was based on the structure of a grid. This was done in order to show if the performance of the kea in choosing the higher amount is different compared to a random distribution of the contained elements in relation to the background. The quantities per stimuli were increased up to the combination 35 vs. 36 as to evaluate the proficiency of kea in this specific set-up, being in the range from 25 vs. 21, 22, 23, 24 up to 36 vs. 32, 33, 34, 35. This has led to 32 combinations, with each of the 11 birds having 10 trials per combination, summing up to 110 trials per combination, 320 trials per bird and 3520 trials in total for all birds and combinations. The kea were tested individually 32 trials per session with a maximum of 2 sessions per day, once antemeridian and once in the afternoon. A fully randomised order of combinations was provided, and due to the architecture of the software, the rewarded side was also fully randomized. Both of the two stimuli versions were utilized for an equal part.
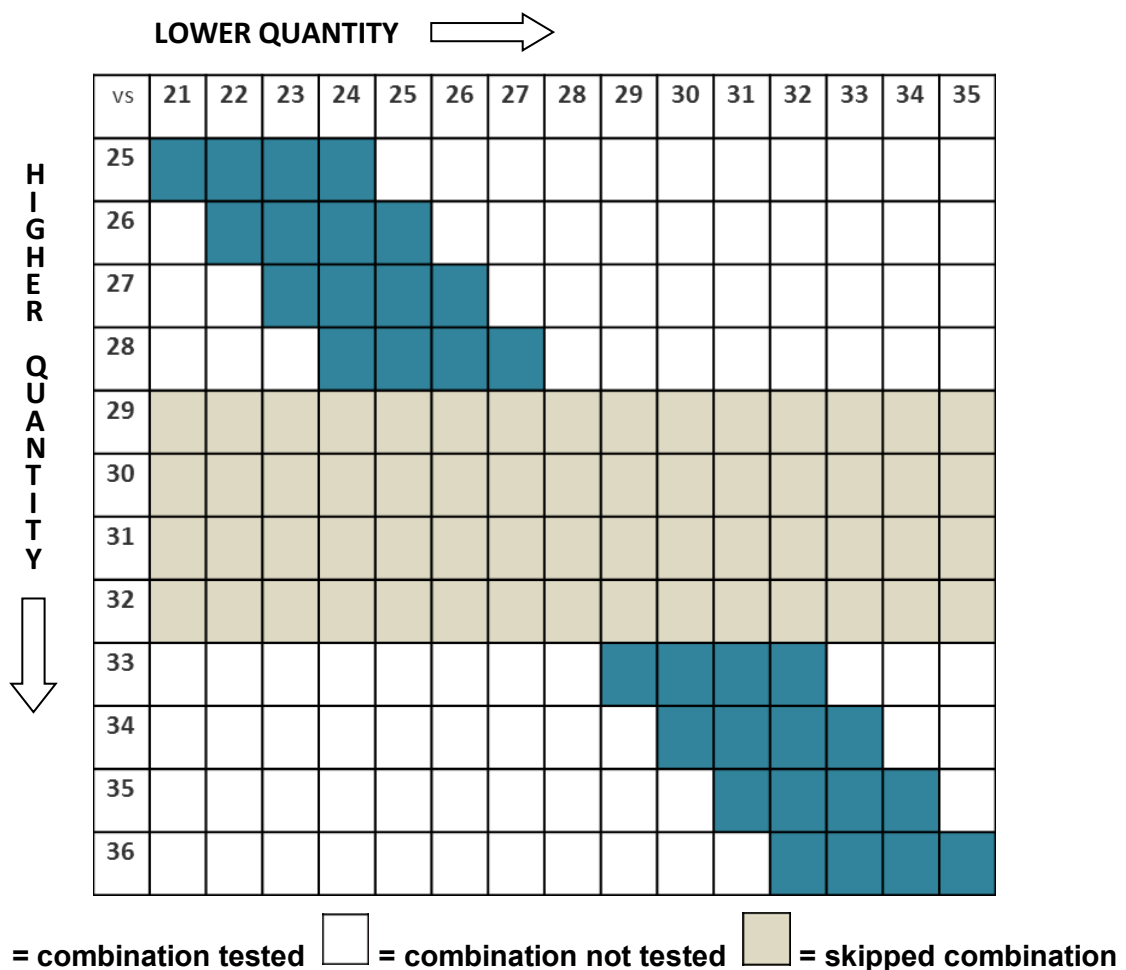
LOWER QUANTITY ⟹

HIGHER QUANTITY ⟱

vs	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
25	■	■	■	■											
26		■	■	■	■										
27			■	■	■	■									
28				■	■	■	■								
29	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
30	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
31	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
32	□	□	□	□	□	□	□	□	□	□	□	□			
33									■	■	■	■			
34										■	■	■	■		
35											■	■	■	■	
36												■	■	■	■

■ = combination tested ☐ = combination not tested □ = skipped combination

Fig 14. Quantity discrimination pairs of stimuli used in Experiment II.

2.3.4. Data Collection and Analysis

2.3.4.1. Data collection (Experiment I)

The data collection for Experiment I was conducted between March the 23rd and May the 18th 2015 at the research Station Haidlhof. The test compartment "*porticula res*" has been used, as it provided more space compared to the alternative compartment, and allowed for a visual separation of the main aviary.

2.3.4.2. Data collection (Experiment I stage II and Experiment II)

The data collection for Experiment I part II as well as for Experiment II was done between May the 21st and June the 25th 2015 also at the Research Station Haidlhof, but utilizing the test compartment "*porticula tabula*" at the opposite side of the station. This was a necessity, because only this compartment provided the use of the touchscreen system.

2.3.4.3. Statistical analysis

The collected data was analysed in utilizing generalized linear mixed models (GLMM) to evaluate if the absolute difference and the relative difference were significant predictors of success rate. Also, three Wald-$\chi 2$-tests were used to compare the overall success rate (i.e., the number of correct trials/number of total trials using the pooled data of all combinations and individuals as well as only the lower quantity combinations and only the higher quantity combinations). As multiple records represent repeated measurements for a single subject, all the data for the GLMMs was structured to be nested in each individual. Thus, for all models, subject identity was entered as random variable with intercept being excluded. A linear model was used for target distribution, and the target was set as the percentage of correct trials. The target variable was tested against the fixed effects of the relative and absolute difference between the two stimuli presented in each trial. Maximum iterations were set at 1000, in order for the algorithm to search for a convergent solution. This was true for the analysis of the data regarding all the experiments conducted. Models including all main effects and several reduced models were run. The best fitting model was selected by comparison of the corrected Akaike Information Criteria (cAIC). In order for the thesis to be concise, only the best fitting models are reported. Statistical data was prepared using Microsoft Excel (version 14.0.4734.1000 [32-bit], part of Microsoft Office Professional Plus 2010, Microsoft Corporation, Redmond, Washington, USA) and statistical tests were carried out with SPSS statistical software (version 20.0, IBM, Armonk, USA). All reported p-values are two-tailed and the significance threshold was fixed at $\alpha \leq 0.05$.

3. Results

3.1. Experiment I results

3.1.1. Experiment I control for the effect of area

In order to control that the birds did not base their choices on the overall area that the elements of a stimulus took up, two versions of each stimulus displaying a quantity were constructed. On one version the elements were closer together and on the other further apart. If the birds chose correctly even though the elements on the correct stimuli were closer together than the ones on the incorrect stimuli, this would prove that they did not rely on the surface area that the elements took up.

A one-way between subjects ANOVA was conducted to compare the effect of area size – meaning the space that the elements took up on the entirety of the stimuli – on the ability of the kea to differentiate between higher amounts and lower amounts. The conditions were that the elements on the stimuli have been closer together either on the side that displayed a higher amount or on the stimuli that displayed a lower amount. As measure of central tendency the percentage of correct trials per combination, regarding one of the conditions, was used. Using Levene Statistic ($p< .048$) as test of homogeneity of variances, it has been shown that the variances were not equal within sample variation, but statistically different from each other. ANOVA generally assumes homogeneity of variance, but it has been proven to be robust (meaning that the alpha value is kept stable between 0.04 and 0.06) against these assumptions, as long as the sample size is large enough (Schmider et al., 2010). As the sample size was 320 for each of the two groups, thereby guaranteeing an alpha level less than 0.06, the robustness of a one-way between groups ANOVA was enough to still yield results. The ANOVA ($p< .001$) showed that there has been a statistically significant difference in the means of the groups. As robust test for the equality of means, the Welch test ($p< .001$) has been used, affirming the effect of the ANOVA.

There was a significant effect of choosing the stimuli on which the higher amount was displayed with elements closer together (~ 1-2 centimetre) over the stimuli displaying a lower amount with the elements further apart (~ 3-4 centimetres) at the $p < .05$ level for the conditions [$F (1, 638) = 56.910$, $p < 0.001$]. These results suggest that the area of the elements of the stimuli does not have an effect on the kea choosing the stimuli displaying the larger amount.

Results

3.1.2. Experiment I control for side bias

The results of the side bias analysis as the relative frequency of left side choices for Experiment I are that the five subjects in sum chose the left side in 1668 of 3200 cases, making up 52.1%.
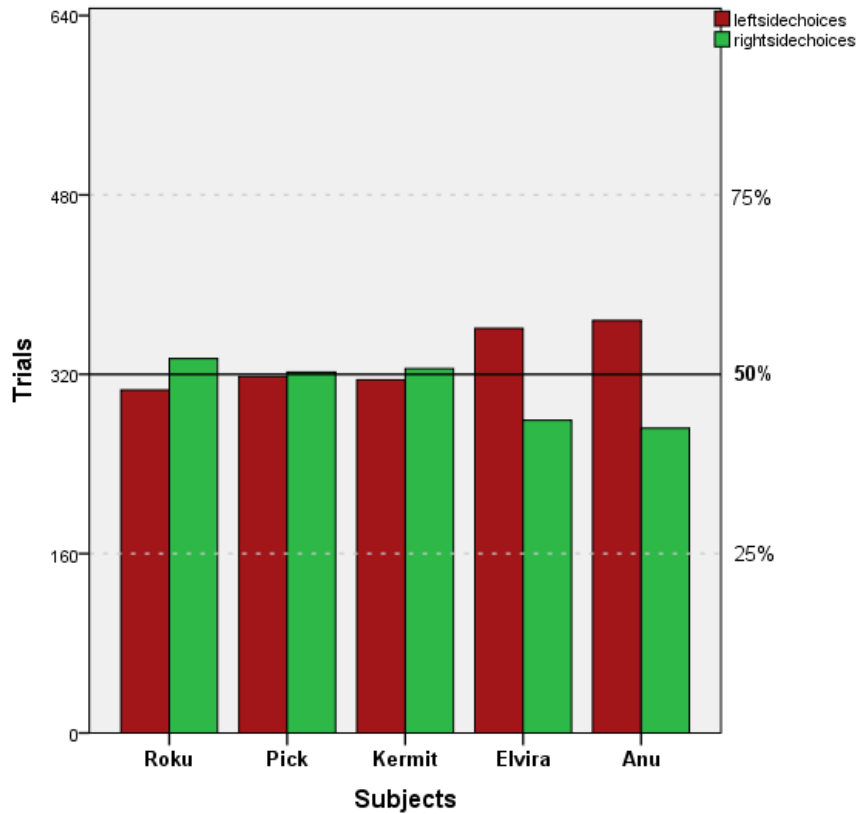

Fig 15. Individual side bias Exp I. Percentage of left side choices and right side
   choices of each individual subject in Experiment I. The solid line represents
   the chance level (50%).

Individually, the left side choices ranged from 47.8% to 57.5% (Roku: 47.8%, Pick: 49.7%, Kermit: 49.2%, Elvira: 56.4%, and Anu: 57.5%), as shown in the previous figure. Because none of the birds chose one side preferably, no subject had to be excluded due to side bias.

3.1.3. Experiment I correct choices compared to chance level

To test if the birds were choosing the stimulus displaying the higher quantity over chance level, three Wald-χ^2-tests were run using a GLM, processing either all the data, only the lower quantity set (combinations from 5 vs. 6 up to 12 vs. 16) or only the higher quantity set (combinations from 13 vs. 14 up to 20 vs. 24). Over all birds and trials including all quantity sets, the kea have been significant above chance level with Wald-χ^2 (1) = 255.835, p < 0.001. The descriptive statistics were as follows:

Results

	N	%
Correct trials	2059	64,3
False trials	1141	35,7
Overall trials	3200	100

Over all birds, but only the low quantity set (combinations from 5 vs. 6 up to 12 vs. 16), they have performed correct significant above chance level with Wald-χ2 (1) = 73.630, p < 0.001. The descriptive statistics are as follows:

	N	%
Correct trials	973	60,8
False trials	627	39,2
Overall trials	1600	100

Over all birds and only the high quantity set (combinations from 13 vs. 14 up to 20 vs. 24), the kea have performed correct significant above chance level with Wald-χ2 (1) = 195,216, p < 0.001. The descriptive statistics are as follows:

	N	%
Correct trials	1086	67,9
False trials	514	32,1
Overall trials	1600	100

We have to be careful with these results however, as the birds in sum reach significance even though the rate of success is not exceeding chance level by far. For this reason, we tested the rate of success of the individual birds by utilizing a binomial test. The following figure demonstrates the percentage of correct choices made in sum by all five subjects regarding specific quantity pairs tested in Experiment I. The figures for the results of all quantity combinations tested in Experiment I can be seen in the Appendix (*Appendix, I.III.I. & I.III.II.*). The line represents the chance level (50%). Asterisks indicate significant deviation from chance level at the individual level: *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001. The number next to the Asterisk represents the number of subjects that deviate significantly from chance level.
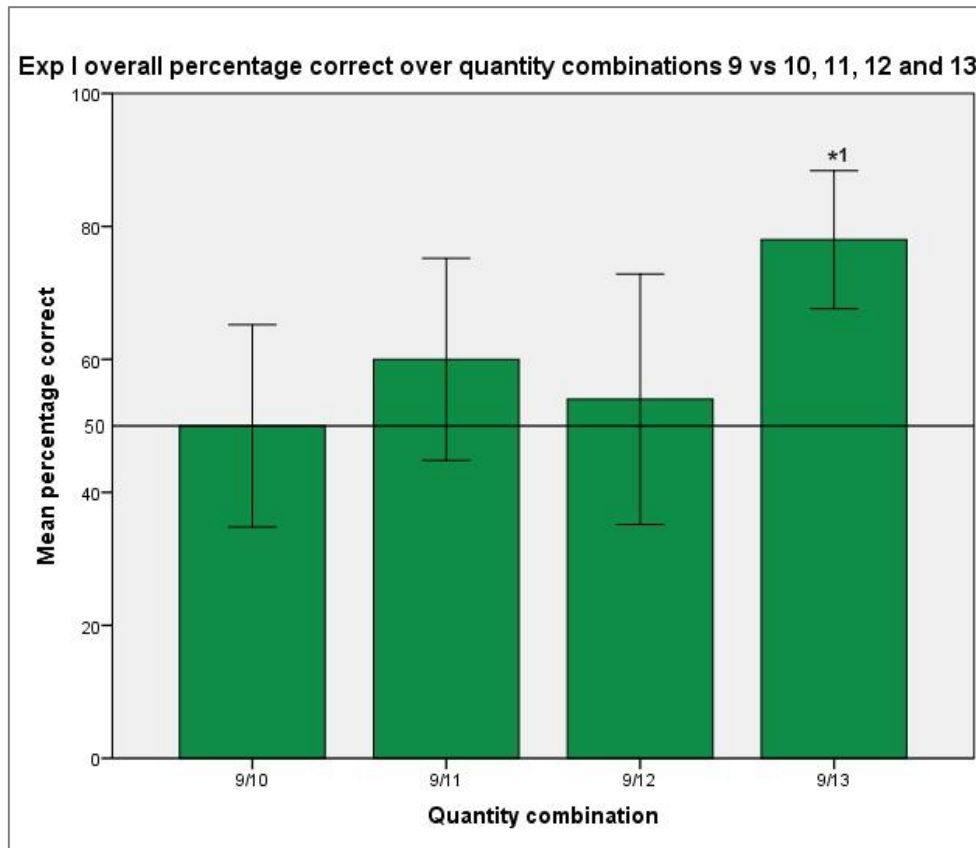
Fig 16. Percentage of correct choices made in sum by all subjects in specific combinations tested in Experiment I.

The table in the Appendix (*Appendix, I.III.III.*) shows the statistical output of the GLM conducted on the group level as well as the outcomes of the binomial testing on the individual level, including the significances per combination tested for Experiment I.

3.1.4. Experiment I correct choices compared to specific probabilities

The binomial testing of the data shows, that in order for a subject to have significant results ($p \leq 0.05$) for a specific combination, they have to choose the higher quantity nine out of ten trials correctly. This has been accomplished for 26 of the 64 combinations tested, by at least one subject and at most by four subjects for a single combination (see *Appendix, I.III.III.*).

Compared to the probability of correctly choosing nine out of ten by chance ($p \leq 0.01$), binomial testing showed that four of the five subjects reached significance for specific combinations (Subject 1 combinations 10 vs. 13 and 14 vs. 18; Subject 2 combinations 13 vs. 17, 14 vs. 18, 17 vs. 19, 20 vs. 22, 20 vs. 23, 20 vs. 24; Subject 3 combinations 5 vs. 8 and 11 vs. 15, Subject 5 combination 14 vs. 18).

When comparing the mean percentage of all subjects trials yielding significant results (test proportion = 0.09%) with the individual subjects percentage of significant outcomes, only one subject reaches significance (Subject 3: p = 0,027). When applying Bonferroni correction, in order to counteract the problem of multiple comparisons, the probability of seeing the observed difference is adjusted (p ≤ 0.001). This correction lowers the family wise error rate and when applied, none of the subjects reach significance in the binomial test.

3.1.5. Experiment I success in regard to absolute and relative quantity difference of the stimuli presented in a trial

Different generalized linear mixed models (GLMM) were run in order to evaluate if the kea`s success in choosing the higher quantity displayed by the two "real-life" stimuli presented in a quantity discrimination trial can be predicted by referring to the absolute or to the relative difference between them. When analysing all quantity combination pairs tested (from 5 vs. 6, 7, 8, 9 up to 20 vs. 21, 22, 23, 24), the best fit model showed that the absolute difference was significant in predicting success in choosing the higher quantity displayed, between the two stimuli presented (GLMM, $\beta \pm SE$ = -0.023 ± 0.002, $F_{1, 371}$ = 1.901, p = 0.001). Relative difference was also significant in predicting success in choosing the higher quantity displayed, between the two stimuli presented (GLMM, $\beta \pm SE$ = 0.023 ± 0.002, $F_{1, 317}$ = 7.572, p = 0.006).

The bar-charts in the figures below show the means of the percentage of correct choices over all birds in choosing the higher amount between quantities of the stimuli pairs tested in relation to the absolute difference and the relative difference, respectively. The line represents the chance level (50%). The error bars indicate individual performance of the subjects.
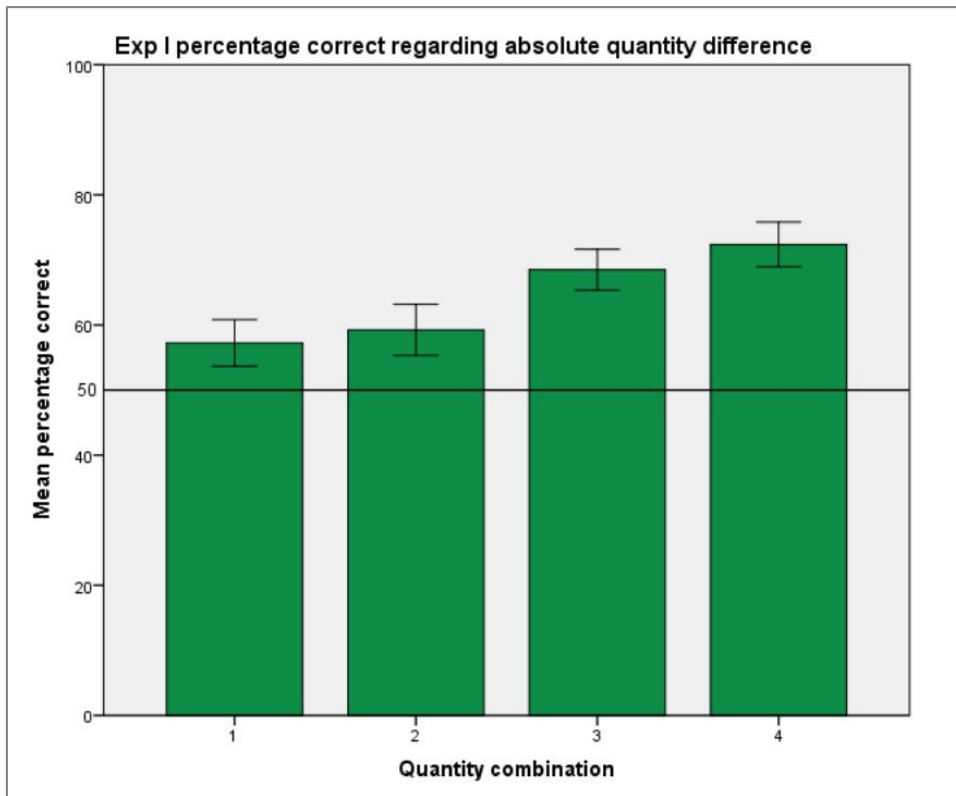
Fig 17. Percentage of correct choices made in sum by all subjects in regard to the absolute differences tested in Experiment I.
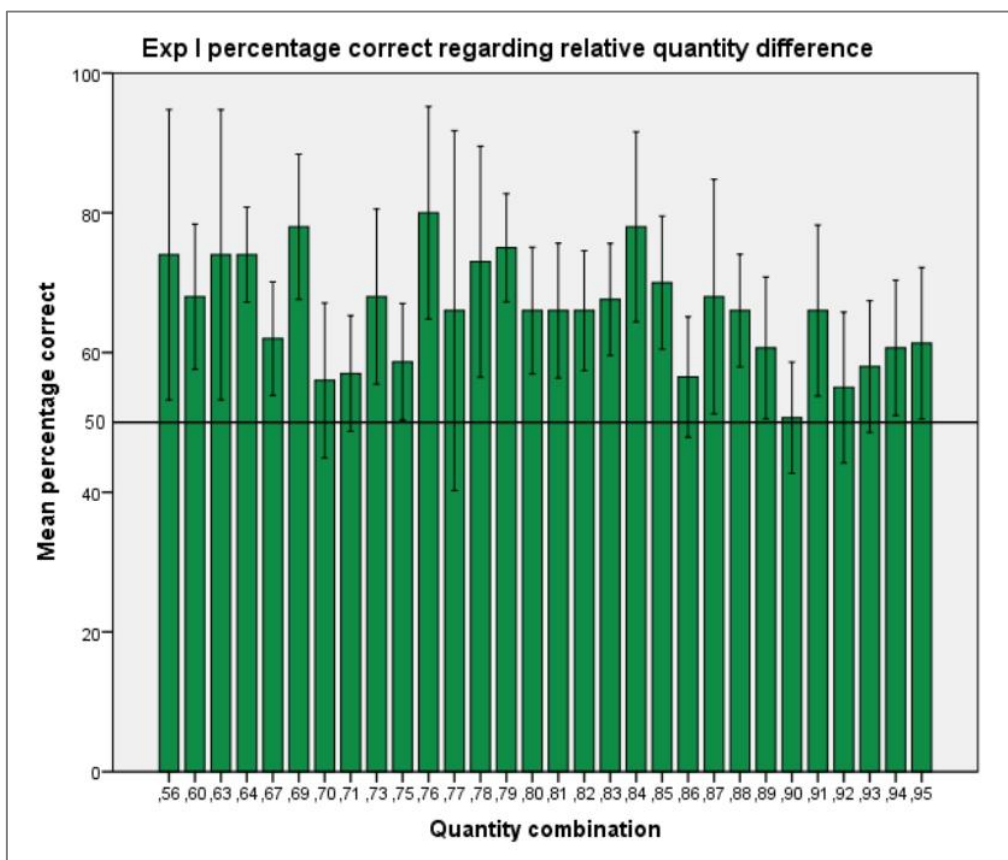

Fig 18. Percentage of correct choices made in sum by all subjects in regard to the relative differences tested in Experiment I.

3.2. Experiment I stage II results

Illustrated in the following figure are the results of the side bias analysis as the relative frequency of left side choices for experiment I part II. In sum, the thirteen subjects chose the left side in 1040 of 2600 cases, making up 40%.
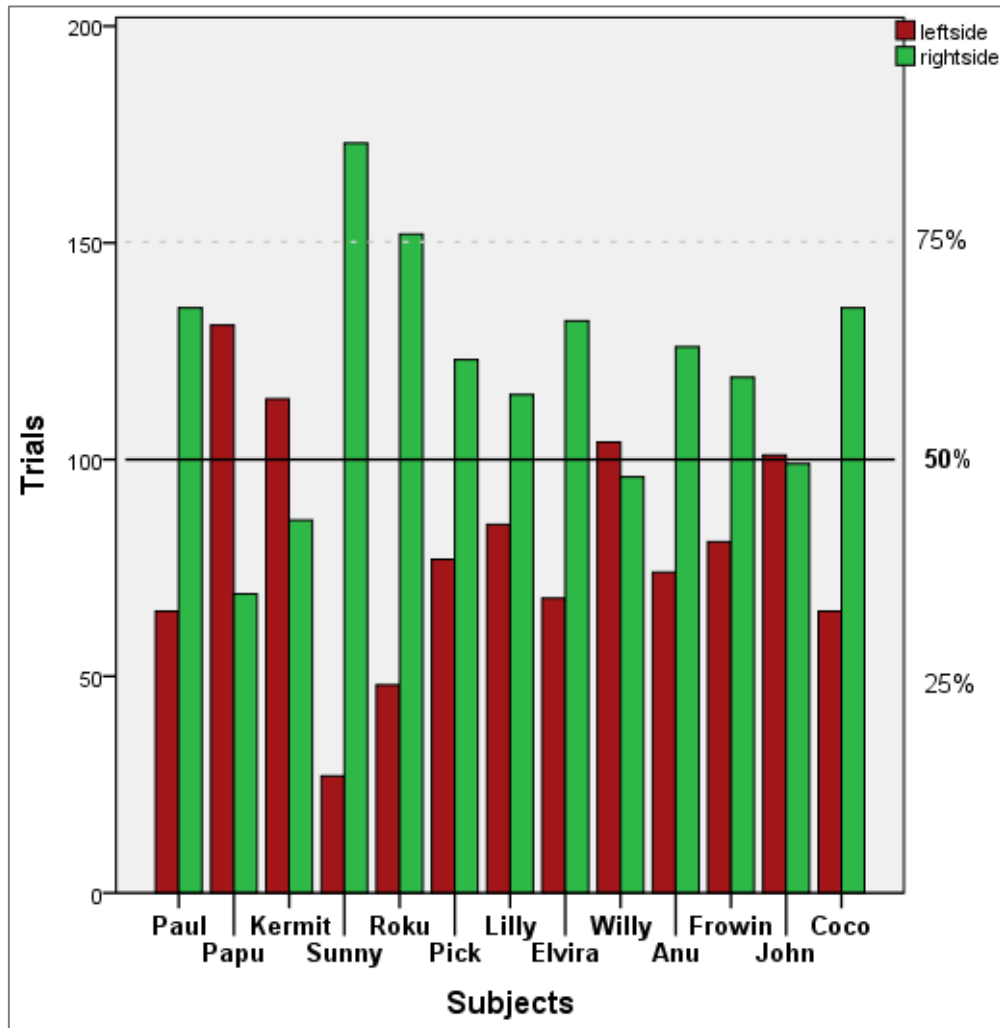
Figure 19. Individual side bias Exp. I stage II. Percentage of the left side and right side choices of each individual subject in Exp I Stage II. The solid line represents the chance level (50%) and the upper dashed line 75%.

The correct left side choices summed up to 600, making up 57.7%, having been a significant proportion of left side choices (Wald-χ2 (1) = 24,419, p < 0.001). Individually, the left side choices ranged from 13.5% to 65.5% (Paul: 32.5%, Papu: 65.5%, Kermit: 57%, Sunny: 13.5%, Roku: 24%, Pick: 38.5%, Lilly: 42.5%, Elvira: 34%, Willy: 52%, Anu: 37%, Frowin: 40.5%, John: 50.5% and Coco: 32.5%). Only one of the birds chose one side exceeding 85%, namely Sunny (86.5% right side choices). As the proportion of correct right side

choices of the overall right side choices was a statistically non-significant contributor, Sunny had to be excluded from the experiment (53,2% correct right side choices, Wald-χ2 (1) = 0.698, p = 0.403), as is visualized in the graph below.
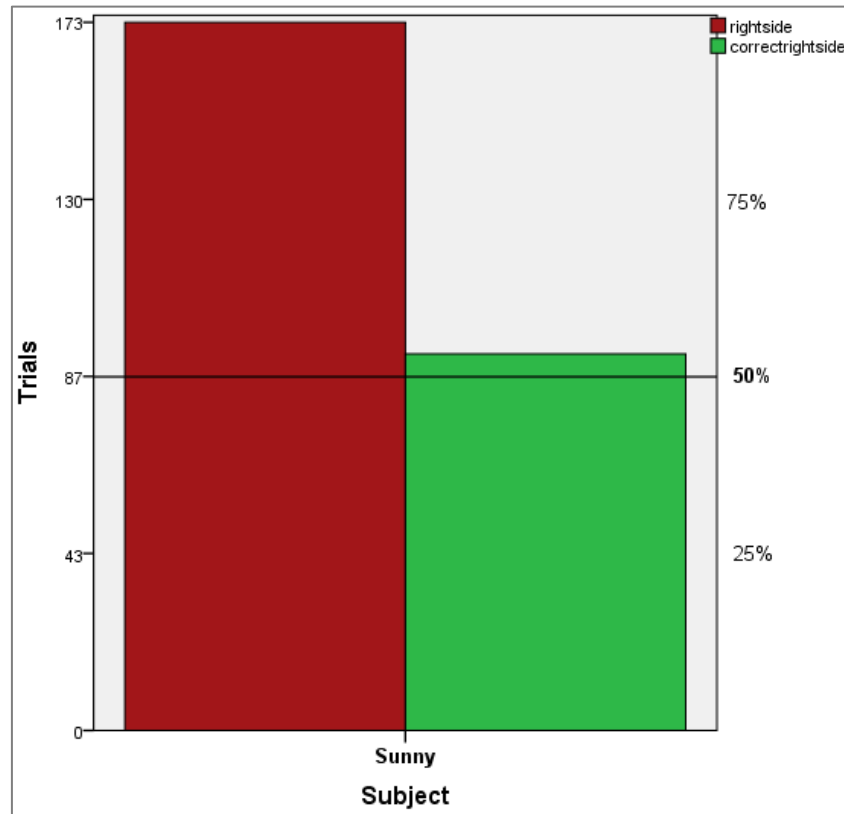

Fig 20. Correct right side choices in relation to right side choices of Sunny in Exp. I stage II. Proportional percentage of the correct right side choices of subject 4 ("Sunny") in Exp. I stage II.

3.2.1. Experiment I stage II correct choices compared to chance level

To test if the birds were choosing the stimulus displaying the higher quantity over chance level during the second stage of Experiment I, a Wald-χ^2-test was run using a GLM, processing either all the data, only the lower quantity set or only the higher quantity set. Over all birds and trials including all quantity sets, the kea have been significant above chance level with Wald-χ^2 (1) = 51.178, p < 0.001. The descriptive statistics are as follows:

	N	%
Correct trials	1483	57,0
False trials	1117	43,0
Overall trials	2600	100

Results

We have to be careful with these results however, as the birds in sum reach significance even though the rate of success is only slightly above chance level. For this reason, we tested the rate of success of the individual birds by utilizing a binomial test. The following figure demonstrates the percentage of correct choices made in sum by all five subjects regarding specific quantity pairs tested in part II of Experiment I. The figures for the results of all quantity combinations tested in Experiment I part II can be seen in the Appendix (Appendix, I.IV.I. & I.IV.II.). The line represents the chance level (50%). Asterisks indicate significant deviation from chance level at the individual level: *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001. The number next to the Asterisk represents the number of subjects that deviate significantly from chance level.
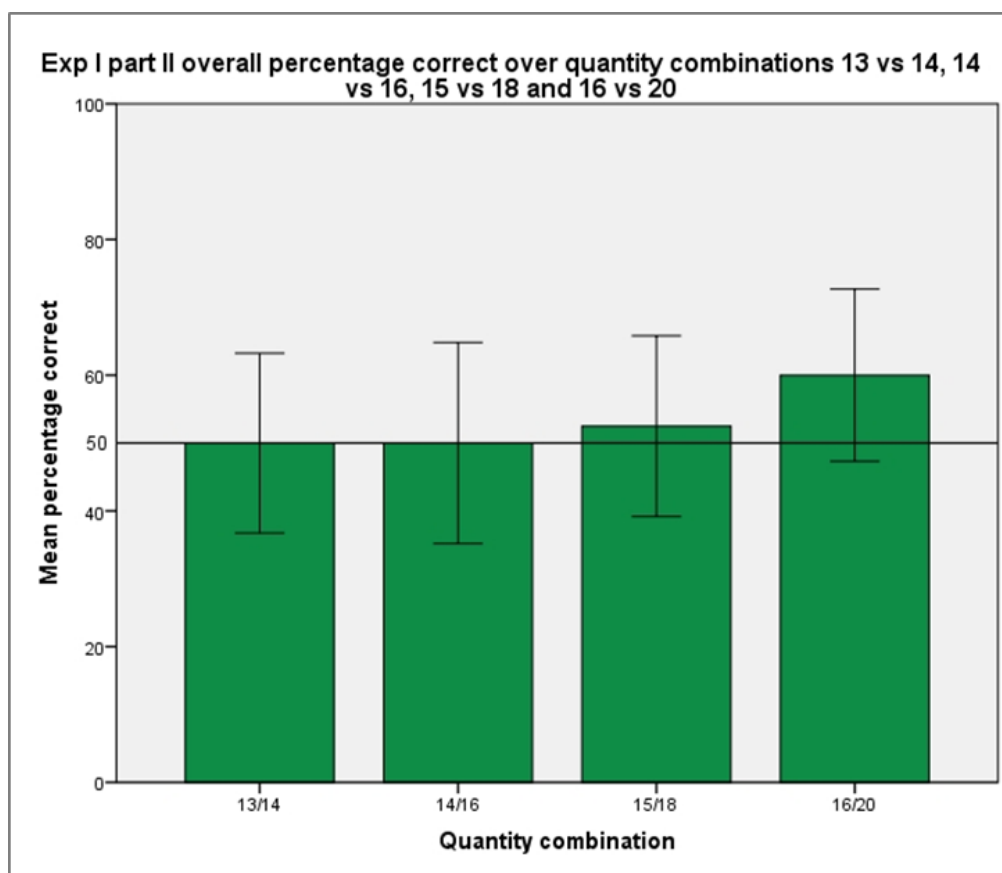
Fig 21. Percentage of correct choices made in sum by all subjects in specific combinations tested in Experiment I part II.

The next figure demonstrates the means of the percentage of correct choices over all birds in choosing the higher amount between quantities in relation to their absolute difference and their relative difference, respectively. The line represents the chance level (50%). Asterisks indicate significant deviation from chance level: *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001. In the Appendix are all the tables regarding the results for the combinations with an absolute difference of one to four (*Appendix I.IV.II.*).
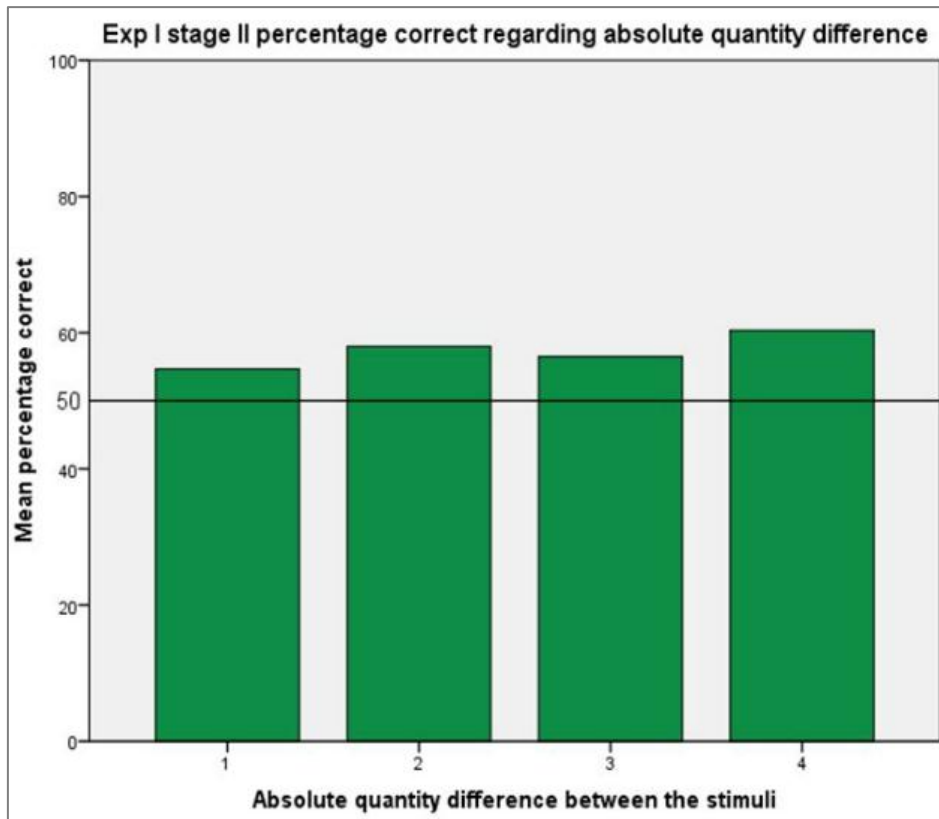
Fig 22. Percentage of correct choices made in sum by all subjects in regard to the absolute quantity differences tested in Experiment I, part II.
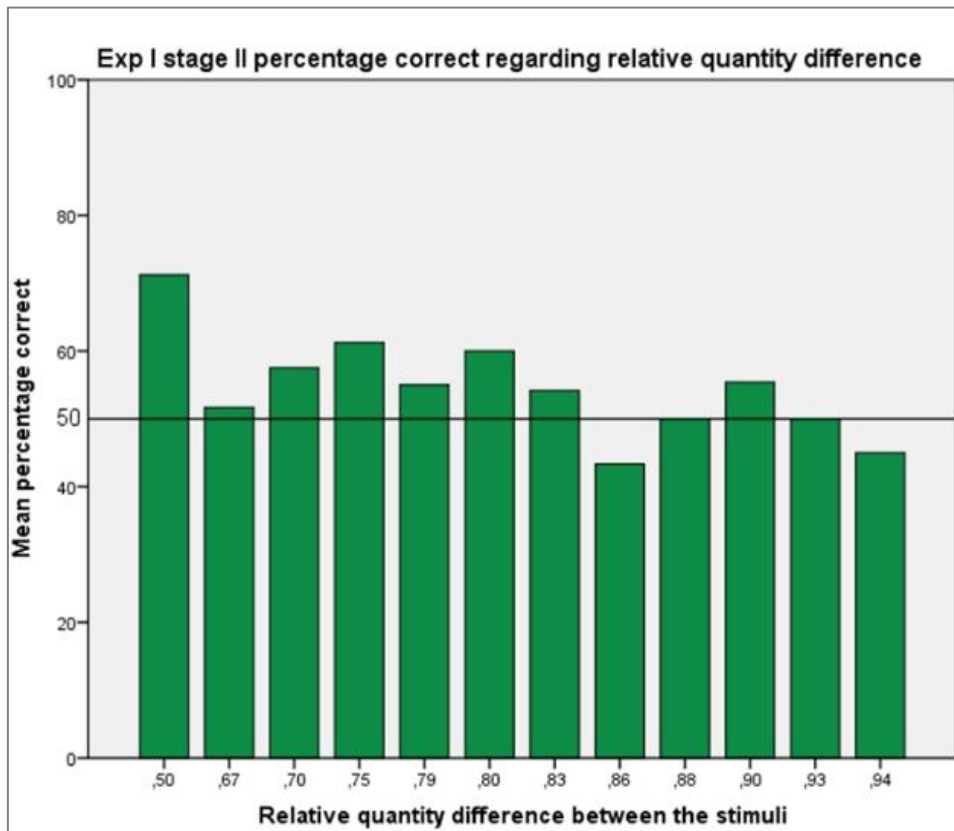

Fig 23. Percentage of correct choices made in sum by all subjects in regard to the relative quantity differences tested in Experiment I part II.

The table in the Appendix (*Appendix, I.IV.III.*) shows the statistical output of the GLM conducted on the group level as well as the outcomes of the binomial testing on the individual level including the significances per combination tested for Experiment I part II.

3.2.2. Experiment I stage II correct choices compared to specific probabilities

The binomial testing of the data shows, that in order for a subject to have significant results (p ≤ 0.05) for a specific combination, they have to choose the higher quantity nine out of ten trials. This has been accomplished by six out of thirteen subjects, for five of the 20 combinations tested (see *Appendix, I.IV.III.*). Compared to the probability of correctly choosing nine out of ten by chance (p ≤ 0.01), binomial testing showed that only one of the thirteen subjects reached significance (Subject 12 combination 4 vs. 8).

When comparing the mean percentage of all subjects trials yielding significant results (test proportion = 0.28%) with the individual subjects percentage of significant outcomes, none of subjects reached significance (closest result by Subject 12: p = 0.053). When applying Bonferroni correction, in order to counteract the problem of multiple comparisons, the probability of seeing the observed difference is adjusted (p ≤ 0.001). This correction lowers the family wise error rate and when applied, none of the subjects reach significance in this binomial test.

3.2.3. Experiment I stage II success in regard to absolute and relative quantity difference of the stimuli presented in a trial

In order to evaluate if the kea`s success in choosing the higher quantity displayed by the two virtual stimuli presented in a quantity discrimination trial can be predicted by referring to the absolute or to the relative difference between them, different generalized linear mixed models (GLMM) were run. The best fitting model with the lowest Akaike Information Criterion (AIC= -226.644) showed that the relative difference was significant in predicting success in choosing the higher quantity displayed, between the two stimuli presented (GLMM, $\beta \pm$ SE = 0.021 ± 0.002, $F_{1, 237}$ = 56.509, p = 0.001). The absolute difference between the quantities was not significant in prediction (GLMM, $\beta \pm$ SE = 0.021 ± 0.002, $F_{1, 237}$ = 0.102, p = 0.750). So, in stage II of Experiment I, relative difference was positively correlated with success of the kea in solving the quantity discrimination task, but absolute difference was not.

3.2.4. Comparison between Experiment I stage I and stage II

A one-way between subjects ANOVA was conducted to compare the effect of the stimuli being "real life" or "virtual" on the ability of the kea to differentiate between stimuli with a higher amount of elements and lower amounts. As measure of central tendency the percentage of correct trials per combination, regarding one of the conditions, was used. The ANOVA ($p < .001$) showed that there has been a statistically significant difference in the means of the groups. Testing with "real life" stimuli led to a higher success rate (mean percentage = 64.34%) compared to using "virtual" stimuli (mean percentage = 57.37%) at the $p < .05$ level for the conditions [$F (1, 558) = 24.053$, $p < 0.001$].

Using Levene Statistic ($p < .364$) as test of homogeneity of variances it has been shown that the variances (or standard deviations) are statistically equal within the sample variation. As robust test for the equality of means, the Welch ($p < .001$) test has been used and affirms the results produced by the ANOVA. In order to assess an effect size measurement, a generalized linear model (GLM) has been used. The measure of effect size shows that 4.1% of the variability in percentage of correct choices is accounted for by the group membership of the birds, either participating in Experiment I part I or part II.

Taken together, these results suggest that the kea's ability to correctly choose the stimuli of a pair displaying a higher amount of elements, is more accurate in a setting utilizing "real-life" stimuli to differentiate compared to a setting in which they have to choose between "virtual" stimuli displayed on a touchscreen. In the figure below the percentage of correct choices made in regard to quantity pairs tested in Experiment I stage I and stage II can be seen side by side.
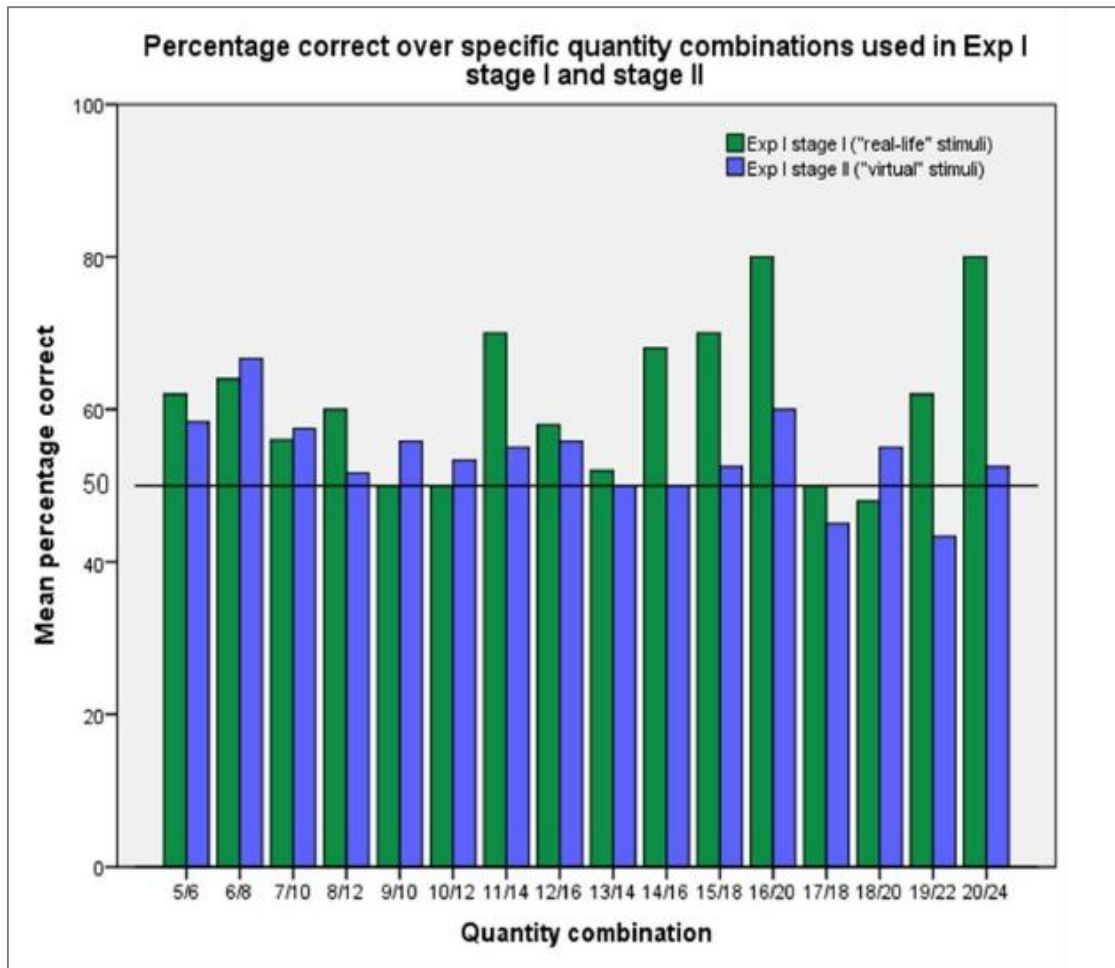
Fig 24. Comparison of correct choices made in sum by all subjects in regard to the quantity combinations tested in Experiment I, part I and part II.

3.3. Experiment II results

3.3.1. Experiment II control for side bias

The results of the side bias analysis as the relative frequency of left side choices for Experiment II are that in sum, the eleven subjects chose the left side in 1289 of 3520 cases, making up 36.6%. The correct left side choices summed up to 698, making up 54.2%, constituting a significant proportion of left side choices (Wald-χ^2 (1) = 8.862, p = 0.003).

Individually, the left side choices ranged between 11.6% and 63.4% (Paul: 57.8%, Kermit: 30.9%, Sunny: 11.6%, Roku: 27.8%, Pick: 48.1%, Lilly: 63.4%, Willy: 34.7%, Anu: 30%, Frowin: 22.5%, John: 58.4% and Coco: 17.5%). The correct left side choices ranged from 43.2% to 60.6% (Paul: 51.3%, Kermit: 60.6%, Sunny: 43.2%, Roku: 56.2%, Pick: 50%, Lilly: 55.2%, Willy: 53.1%, Anu: 52.1%, Frowin: 54.2%, John: 59.3% and Coco: 51.8%), as is shown in the following figure.
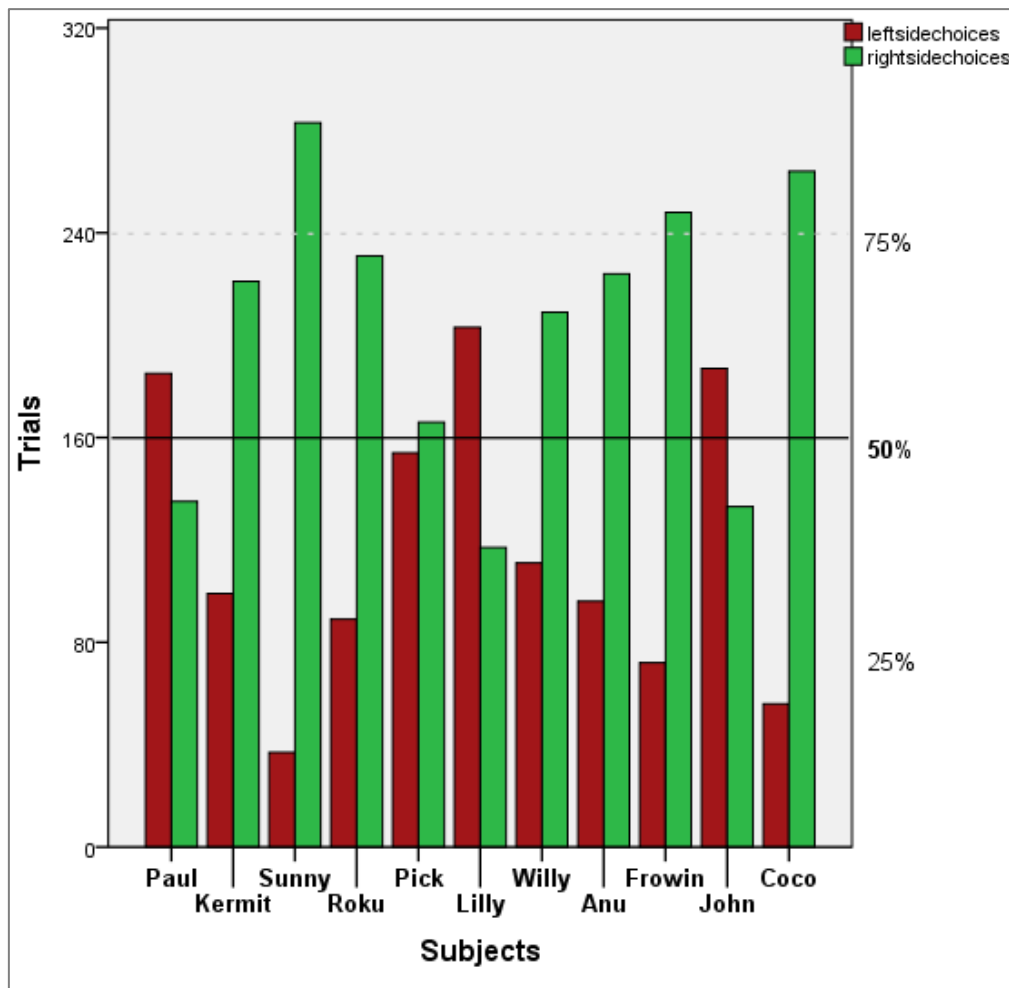
Fig 25. Individual side bias Exp. II. Percentages of left side and right side choices of
each individual subject in Exp. II. The solid line represents the chance level
(50%), the lower dashed line represent 25% and the upper dashed line 75%.

Individually, the left side choices ranged between 11.6% and 63.4% (Paul: 57.8%, Kermit:
30.9%, Sunny: 11.6%, Roku: 27.8%, Pick: 48.1%, Lilly: 63.4%, Willy: 34.7%, Anu: 30%,
Frowin: 22.5%, John: 58.4% and Coco: 17.5%). The correct left side choices ranged from
43.2% to 60.6% (Paul: 51.3%, Kermit: 60.6%, Sunny: 43.2%, Roku: 56.2%, Pick: 50%,
Lilly: 55.2%, Willy: 53.1%, Anu: 52.1%, Frowin: 54.2%, John: 59.3% and Coco: 51.8%).

Only one of the birds chose one side exceeding 85%, namely Sunny (88.4% right side
choices), and as the proportion of correct right side choices of the overall right side choices
was a statistically non-significant contributor, Sunny had to be excluded from the experiment
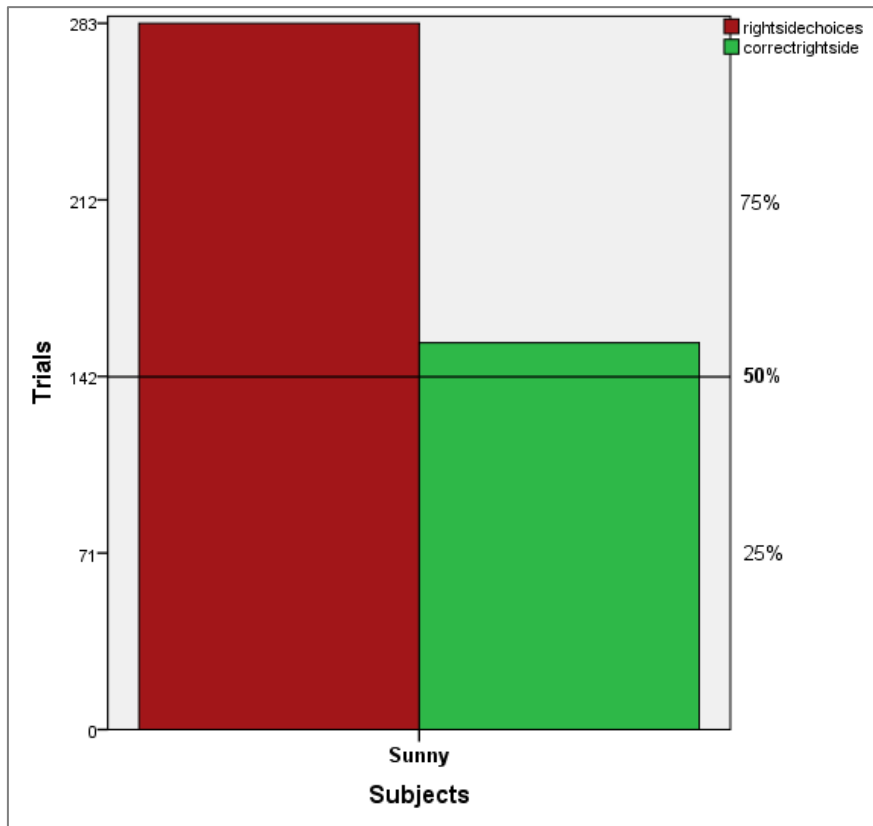(54,8% correct right side choices, Wald-$\chi 2$ (1) = 2.568, p 0.109).

Fig 26. Correct right side choices in relation to overall right side choices of Sunny in Exp. II. Proportional percentage of the correct right side of subject 4 ("Sunny") in Experiment II.

3.3.2. Experiment II correct choices compared to chance level

To test if the birds were choosing the stimulus displaying the higher quantity over chance level in Experiment II, a Wald-χ^2-test was run using a GLM, processing either all the data, only the lower quantity set or only the higher quantity set. Over all birds and trials including all quantity sets, the kea have been significant above chance level with Wald-χ^2 (1) = 24.498, p < 0.001. The descriptive statistics are as follows:

	N	%
Correct trials	1907	54,2
False trials	1613	45,8
Overall trials	3520	100,0

We have to be careful with these results however, as the birds in sum reach significance even though the rate of success is only slightly above chance level. For this reason, we tested the rate of success of the individual birds by utilizing a binomial test. The following

figure demonstrates the percentage of correct choices made in sum by all five subjects regarding specific quantity pairs tested in Experiment II. The figures for the results of all quantity combinations tested in Experiment II can be seen in the Appendix (Appendix, I.V.I. & I.V.II.). The line represents the chance level (50%). Asterisks indicate significant deviation from chance level at the individual level: *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001. The number next to the Asterisk represents the number of subjects that deviate significantly from chance level.
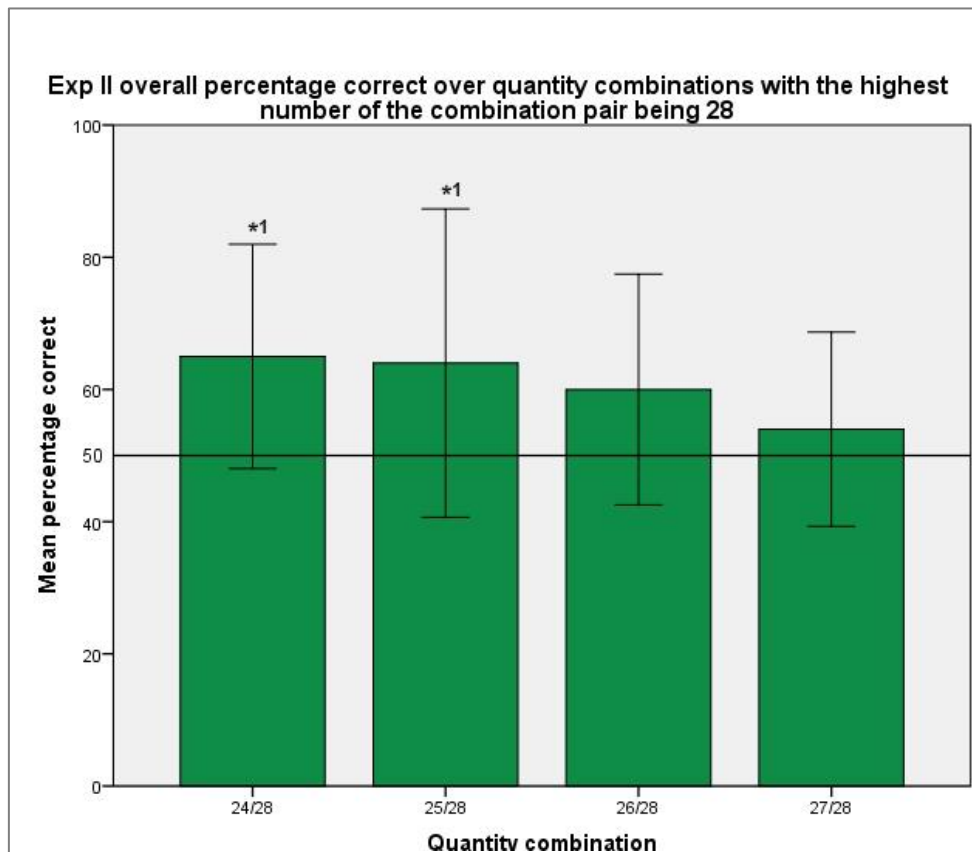

Fig 27. Percentage of correct choices made in sum by all subjects regarding specific combinations used in Experiment II.

The table in the Appendix (Appendix, I.V.III.) shows the statistical output of the GLM conducted on the group level as well as the outcomes of the binomial testing on the individual level, including the significances per combination tested for Experiment II.

3.3.3. Experiment II correct choices compared to specific probabilities

The binomial testing of the data shows, that in order for a subject to have significant results (p ≤ 0.05) for a specific combination, they have to choose the higher quantity nine out of ten trials. This has been accomplished for nine of the 32 combinations tested, by at least one subject (see *Appendix, I.V.III.*). Compared to the probability of correctly choosing nine out of

ten by chance (p ≤ 0.01), binomial testing showed that only one out of the eleven subjects reached significance (Subject 2 combination 25 vs. 27).

When comparing the mean percentage of all subjects trials yielding significant results (test proportion = 0.35%) with the individual subjects percentage of significant outcomes, none of subjects reached significance.

3.3.4. Experiment II success in regard to absolute and relative quantity difference of the stimuli presented in a trial

Also for Experiment II different generalized linear mixed models (GLMM) were run in order to evaluate if the kea`s success in choosing the higher quantity in the quantity discrimination task can be predicted by referring to the absolute or to the relative difference between them. The stimuli pairs in this Experiment were also presented virtually on the VCCT, but the elements of them were placed on fixed symmetrical positions, locked on an invisible grid. When analysing the data received by testing all combinations (25 vs. 21, 22, 23, 24 up to 36 vs. 32, 33, 34, 35), the best fitting model showed that the absolute difference was not significant in predicting success in choosing the higher quantity displayed, between the two stimuli presented (GLMM, $\beta \pm SE = 0.026 \pm 0.002$, $F_{1, 317} = 0.925$, $p = 0.337$). The relative difference was also not significant in prediction (GLMM, $\beta \pm SE = 0.026 \pm 0.002$, $F_{1, 317} = 0.019$, $p = 0.890$). Hence, over all combinations tested in Experiment II, neither absolute nor relative difference was positively correlated with success of the kea in solving the quantity discrimination task.

The figures below show the means of the percentage of correct choices over all birds in choosing the higher amount between quantities of the stimuli pairs tested in relation to the absolute difference and the relative difference, respectively. The line represents the chance level (50%). Asterisks indicate significant deviation from chance level: *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001.
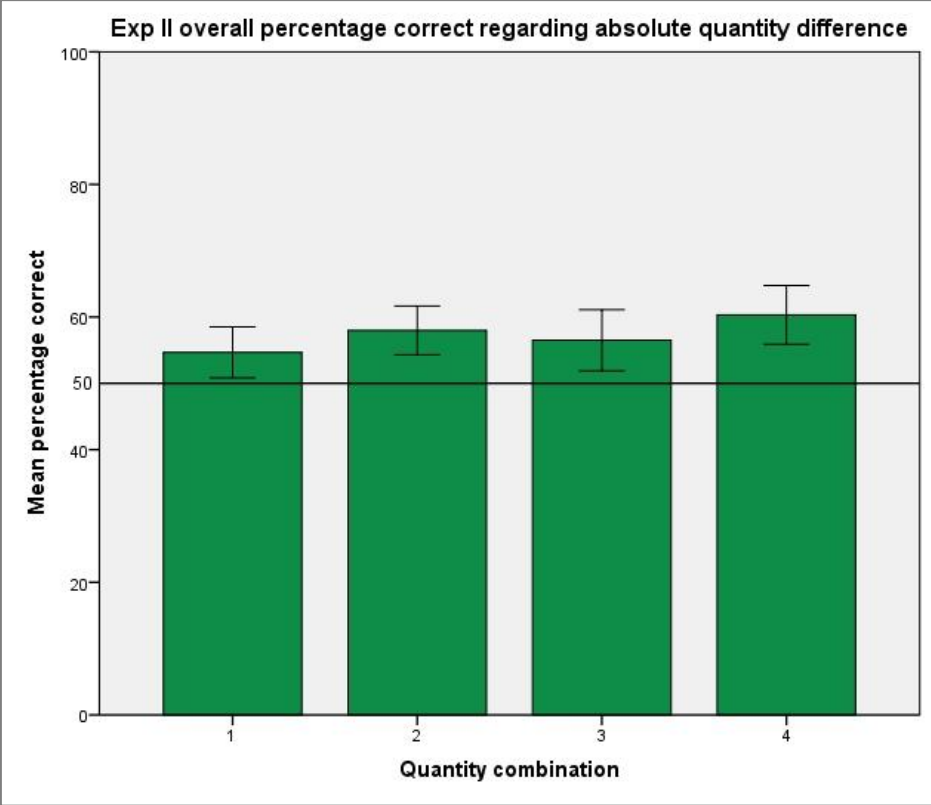
Fig 28. Percentage of correct choices made in sum by all subjects in regard to the absolute differences tested in Experiment II.
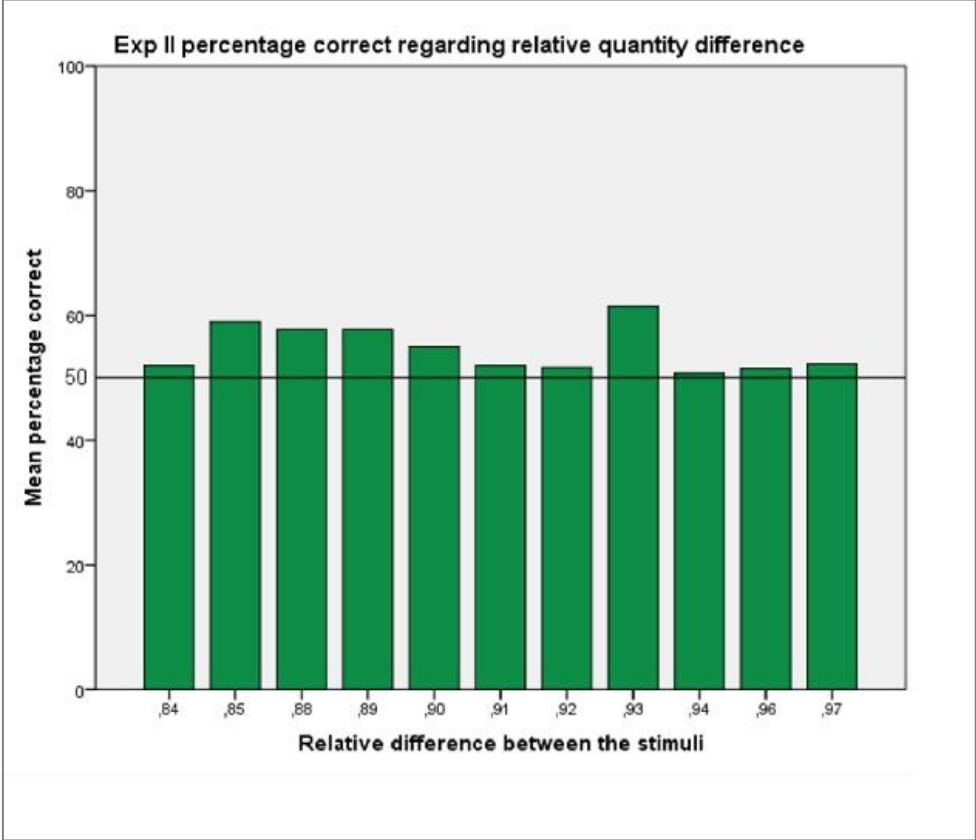

Fig 29. Percentage of correct choices made in sum by all subjects in regard to the relative differences tested in Experiment II.

When analysing the combinations from 25 vs. 21, 22, 23, 24 up to 34 vs. 30, 31, 32, 33, the best fitting model showed that the absolute difference was significant in predicting success in choosing the higher quantity (GLMM, $\beta \pm SE = 0.026 \pm 0.002$, $F_{1, 237} = 5.826$, $p = 0.017$). Contrariwise, the relative difference was not significant in prediction (GLMM, $\beta \pm SE = 0.026 \pm 0.002$, $F_{1, 317} = 0.019$, $p = 0.890$). Hence, when analysing all but the highest 8 combinations (35 vs. 31, 32, 33, 34 and 36 vs. 32, 33, 34, 35) of Experiment II, absolute but not relative difference was positively correlated with success of the kea in solving the quantity discrimination task.

4. Discussion

The kea in sum perform better than chance at choosing the larger quantity of the two stimuli presented as the difference between the amounts relative to the size of the amounts increased. Over all combinations the absolute and the relative difference both were positively correlated with their rate of success in choosing the higher of two presented quantities. Regarding individual subject performance the results of the binomial tests from Experiment I show that four of the five subjects reach significance for nine specific combinations. However, when analysed at the group level, the kea‟s performance did not reach significance. When analysing the distribution of the number of successful outcomes, the performance of the kea does not decrease with increasing overall quantity as predicted by Weber`s law (Gallistel & Gelman, 2005).

The inconsistent results are likely not an outcome of the kea`s visual acuity as colour vision is of high importance for birds foraging and mating behaviours in general (Doucet et al., 2004). Their visual system is highly adapted for colour detection and all bird species of the order Psittaciformes possess ultra-violet sensitive photoreceptors, making them maximally sensitive to wavelengths even below 400 nanometres (Carvalho et al., 2010). The reason more likely could be due to a "Clever Hans"-Effect, meaning that the kea reacted to subtle, unconscious hints given by the experimenter. A way to counteract this would have been to hide the face of the experimenter under a mask, or to place an additional sight barrier in front of the bird while setting up each trial.

The results of Experiment I stage II suggest that the ability to correctly choose the stimulus of a pair displaying a higher amount of elements is more accurate in a setting utilizing "real-life" stimuli compared to a setting in which they have to choose between "virtual" stimuli displayed on a touchscreen. The statistical model showed that relative difference was positively correlated with success of the kea in solving the quantity discrimination task, but absolute difference was not. So – in contrast to part I of Experiment I – in part II of Experiment I the kea are relying on the relative difference between quantities. This makes it possible that there is a diversion in cognitive mechanisms underlying the keas approaches in solving the task compared to the experiment using real-life objects. Regarding individual subject performance, the results of the binomial tests from Experiment I stage II show, that six of the thirteen subjects reached significance for five specific combinations. However, when analysed at the group level, the kea‟s performance did not reach significance. When analysing the distribution of the number of successful outcomes, the performance of the kea does decrease with increasing overall quantity as predicted by Weber`s law. As part II of

Discussion

Experiment I was conducted using the touchscreen, these results suggest that the above assumption regarding a "Clever Hans"-Effect distorting the outcome in the first part of the experiment is possible.

In Experiment II it was shown that when the organization of the graphic elements of two "virtual" stimuli displayed on a touchscreen is based on the structure of a grid, the performance of kea in choosing the higher amount is significantly above chance level. This holds true even for higher quantity combinations compared to when the elements of the two stimuli were randomly distributed in stage II of Experiment I. When leaving out the highest eight combinations tested in the analysis of Experiment II, absolute quantity difference is the significant reference point for the birds in making their choice. This could imply that the kea have been relying on different problem solving mechanisms in the lower range combinations tested, as they used the absolute quantity difference as indicator. It is possible that the birds used "subitizing" in the lower ranges, then switching to "simple estimation" as the range of combination pairs increased during testing (Davis & Pérusse, 1988). The reason why they rely on absolute difference to a higher degree when the elements of the stimuli are placed according to a grid could be accounted for by the symmetrical presentation, since it visualizes the difference in quantity more pronouncedly. In the study of human psychology by Jansen and colleagues it was argued that children could use the same cognitive mechanism of subitizing even for larger quantities, if the stimuli elements are arranged according to a symmetrical pattern (Jansen et al., 2014). This would mean that the kea have a higher threshold for using "subitizing" than is assumed for other species in the literature (Taves, 1941; Mandler & Shebo, 1982; Dehaene, 1992; Anobile et al., 2015). Even though the birds had to choose the one stimulus displaying a higher number of elements, they could have avoided the one with the grid missing more elements instead. This way of solving the task is more effective because a comparably much smaller amount of missing elements is giving the same information about the displayed quantity as estimating all the displayed elements. It would require more research in order to determine which mechanisms the birds utilize to make their decision. Regarding individual subject performance the results of the binomial tests from Experiment II show, that four of the eleven subjects reached significance for five specific combinations. However, when analysed at the group level, the kea"s performance did not reach significance. When analysing the distribution of the number of successful outcomes, the performance of the kea does decrease with increasing overall quantity as predicted by Weber`s law but also increases in the middle section of the quantity range. This increase could be due to the kea adjusting their approach to solving the problem, possibly by relying on missing elements as suggested above. The fall in success rate for the highest

Discussion

quantities tested could be explained by reaching the upper level limit of their ability for quantity discrimination.

When comparing the test results of this quantity discrimination experiments with that of other species, we have to consider that the experimental setups used are diverse and that the combinations tested are not increasing as to show the maximum capacity for quantity discrimination in all studies. The highest combination of Experiment I, using real life stimuli, correctly solved by a single kea significantly above chance level was 20 vs. 24. The highest combination of Experiment I correctly solved by the most kea above chance level was 14 vs. 18. In part II of Experiment I, using virtual stimuli, the highest combination correctly solved by a single kea significantly above chance level was 4 vs. 8. In comparison, the highest combination that was successfully solved by dogs (*Canis lupus familiaris*) significantly above chance level was 3 vs. 5, failing at 4 vs. 5 (Ward & Smuts, 2007). Wolves (*Canis lupus*) were able to discriminate 4 vs. 5 significantly above chance level, even by having to rely solely on mental representations (Range et al., 2012). Western lowland gorillas (*Gorilla gorilla gorilla*) were able to solve a task with moving stimuli up to the combination 11 vs. 12 successfully, significantly above chance level (Beran et al., 2013). Chimpanzees (*Pan troglodytes*) were able to solve the task up to combination 3 vs. 5 significantly above chance level, but were not tested for combinations above (Vonk et al., 2014). Compared to other bird species, the kea also performed very well. New Zealand robins (*Petroica australis*) successfully solved the combination 4 vs. 6 significantly above chance level but failed to do so for 4 vs. 8 or higher (Hunt et al., 2008). Jungle Crows (*Corvus macrorhynchos*) solved the combination 5 vs. 8 above chance level, whereas Jackdaws (*Corvus monedula*) solved combinations up to 2 vs. 5 (Bogale et al., 2011; Ujfalussy et al., 2013). The highest combination successfully solved by an African grey parrot (*Psittacus erithacus*) significantly above chance level was 3 vs. 5, but no combination above (Aïn et al., 2008). Although the kea exceed all the other species in their ability for quantity discrimination, we have to be careful with the comparison, as the results of Experiment I could have been corrupted by a "Clever Hans"-Effect.

In respect to the literature on quantity discrimination theory it can be assumed that the kea did not use the most basic system for making numerosity judgements as the range of quantities tested was beyond the range that *subitizing* is assumed to be used for (Taves, 1941; Mandler & Shebo, 1982; Dehaene, 1992; Anobile et al., 2015). Considering this, they seem to rely on *simple* or *enhanced estimation* (Davis & Pérusse, 1988). As the statistical analysis showed, their success was predicted either by the absolute, the relative or both differences in quantity, depending on the experimental setup. This suggests that different cognitive mechanisms and a combination of them play a role in their ability to discriminate

quantities. Regrettably, the study design has not allowed for an evaluation of the cap in the ability of the kea for quantity discrimination. Due to the stimulus size, it was not reasonable to further increase the amount of elements represented and thus quantity combinations tested. To present a more detailed account of the kea`s sophisticated quantity judgement skills and to understand the cognitive mechanisms used by them in order to accomplish such tasks, further research will be needed.

On a final note, we hope that these results are additional proof for the astonishing cognitive abilities of the kea, and will be referred to by animal scientists and ethicists in order to increase their moral status and to heighten ethical considerations from humans towards these incredible creatures.

5. Abstract

The ability for quantity discrimination in kea (*Nestor notabilis*) has been tested using three different setups. In part I of Experiment I, a two-way choice task with three dimensional stimuli was conducted. The kea had to choose the higher quantity of a pair of stimuli. The absolute difference as well as the relative difference between the combinations was positively correlated with success in solving the task. In part II of Experiment I, a two-way choice task on the VCCT-Touchscreen was conducted. The birds had to choose the higher quantity of a pair of virtual stimuli in the range of the set of combinations provided in part I of Experiment I. The relative difference was significant in predicting success in choosing the higher quantity displayed, but the absolute difference was not. Testing with "real life" stimuli led to a higher success rate compared to using picture stimuli. In Experiment II a two-way choice task on the VCCT-Touchscreen was conducted with the stimuli elements placed on symmetrical positions. The performance of kea was higher, compared to when the elements of the two stimuli were randomly distributed, as in stage II of Experiment I. As the range of quantities tested was beyond the assumed range for *"subitizing"*, we assume that the kea did not utilize this basic system for making numerosity judgements. Instead, they seem to have relied on *"simple"* or *"enhanced estimation"*. As the statistical analysis showed, their success was predicted either by the absolute, the relative or both differences in quantity, depending on the experimental setup. In Experiment I, the performance of the kea does not decrease with increasing overall quantity as predicted by Weber`s law. This could be explained by a "Clever Hans"-Effect, as it was not the case in Experiment I part II, where virtual stimuli were used.

6. Zusammenfassung (Abstract in German)

Die Fähigkeit von neuseeländischen Bergpapageien (Kea) zur Unterscheidung verschieden großer Mengen wurde unter drei Bedingungen untersucht. Im ersten Teil des Experiments I wurde eine Aufgabe mit zwei Entscheidungsmöglichkeiten durchgeführt, in der die Kea von zwei Reizen jenen mit der größeren Menge von Elementen auswählen mussten. Es wurde gezeigt, dass die Kea umso erfolgreicher waren, je größer der Unterschied in der Anzahl der Elemente war, und zwar sowohl absolut als auch relativ gemessen. Im Teil II des ersten Experiments wurde die Aufgabe auf einem berührungsempfindlichen Computermonitor durchgeführt. Die höhere Menge der Stimulus-Paare musste ausgewählt werden, wobei die Mengen im gleichen Bereich lagen, welcher im ersten Teil des Experiments getestet wurde. Hierbei wurde gezeigt, dass die Kea umso erfolgreicher waren, je größer der Unterschied in der Anzahl der Elemente war, jedoch nur relativ gemessen. Auch wurde gezeigt, dass die „lebensechten" Reize im Vergleich zu den abgebildeten Reizen mit höherem Erfolg gewählt wurden. Der Erfolg im ersten Teil des Experiments folgte, im Gegensatz zum zweiten Teil, nicht Weber`s Gesetz, laut welchem ein bemerkbarer Unterschied zwischen zwei Reizen in einem konstanten Verhältnis zur Größe des Bezugsreizes steht. Dies könnte daran liegen, dass die Kea anhand der Gestik oder Mimik des Forschers die richtige Lösung erkannten, da im Teil II des ersten Experiments virtuelle Stimuli verwendet wurden und somit kein menschlicher Einfluss gegeben war. In Experiment II wurde ebenfalls eine Aufgabe mit zwei Entscheidungsmöglichkeiten auf einem berührungsempfindlichen Computermonitor durchgeführt, bei der die Elemente der Reize nun allerdings symmetrisch positioniert waren. Es wurde gezeigt, dass die Kea im Vergleich zum Teil II des ersten Experiments auch in höheren Kombinationen über dem Zufallsniveau lagen. Die Vögel haben dabei möglicherweise die Anzahl der ausgelassenen Elemente verglichen, um zwischen den Mengen zu unterscheiden. Eventuell haben sie auch anhand der geometrischen Form die größere Anzahl erkannt. Diese Fähigkeit wird in der Literatur jedoch nur für niedrige Mengen angenommen, womit Kea im Vergleich zu anderen Spezies herausragen würden.

References

Abramson, J.Z., Hernández-Lloreda, V., Call, J., & Colmenares, F. (2011). Relative quantity judgments in South American sea lions (*Otaria flavescens*). Animal Cognition, vol. **14**, pp. 695–706

Abramson, J.Z., Hernández-Lloreda, V., Call, J., & Colmenares, F. (2013). Relative Quantity Judgments in the Beluga Whale (*Delphinapterus Leucas*) and the Bottlenose Dolphin (*Tursiops Truncatus*). Behavioural Processes, vol. **96**, pp. 11–19

Anobile, G., Cicchini, G.M., & Burr, D.C. (2015). Number As a Primary Perceptual Attribute: A Review. *Perception*, pp. 1–27

Aïn, S.A., Giret N., Grand, M., Kreutzer, M., & Bovet, D. (2008). The discrimination of discrete and continuous amounts in African grey parrots (*Psittacus erithacus*). Animal Cognition, vol. **12**, August: pp. 145–154

Agrillo, C., Piffer, L., Dadda, M., Bisazzo, A., & Serena, G. (2009). Fish can use numerical information when discriminating between small discrete quantities. CogSci 2009 Proceedings, pp. 2414–2419

Agrillo, C., & Beran, M.J. (2013). Number without language: comparative psychology and the evolution of numerical cognition, in Agrillo, Christian/Beran, Michael (ed.): Frontiers in Psychology, vol. **4**, May: pp. 6–8

Beauchamp, G. (2003). Group-size effects on vigilance: a search for mechanisms. Behavioural Processes, vol. **63**, pp. 111–121

Beran, M.J. (2001). Summation and numerousness judgments of sequentially presented sets of items by chimpanzees (*Pan troglodytes*). Journal of Comparative Psychology, vol. **115**, pp. 181–191

Beran, M.J., McIntyre, J.M., Garland, A., & Evans, T.A. (2013). What counts for „counting‟? Chimpanzees, *Pan troglodytes*, respond appropriately to relevant and irrelevant information in a quantity judgment task. Animal Behaviour, vol. **85**, no. 5, pp. 987–993

Beran, M.J., Parrish, A.E., & Evans, T.A. (2015). Numerical Cognition and Quantitative Abilities in Nonhuman Primates, in: David C. Geary/Daniel B. Berch/Kathleen M. Koepke (eds.), Evolutionary Origins and Early Development of Number Processing Volume 1, London, San Diego, Waltham, Oxford: Elsevier Inc., pp. 91–112

References

Bogale, B.A., Kamata, N., Mioko, K., & Sugita, S. (2011). Quantity discrimination in jungle crows, *Corvus macrorhynchos*. Animal Behaviour, vol. **82**, pp. 635–641

Boysen, S.T., & Berntson, G.G. (1989). Numerical Competence in a Chimpanzee (*Pan troglodytes*). Journal of Comparative Psychology, vol. **103**, no. 1, pp. 23–31

Brejaart, R. (1988). Diet and Feeding Behaviour of the Kea (*Nestor notabilis*). Lincoln University Digital Dissertation.

Cantlon, J.F., & Brannon, E.M. (2010). Animal Arithmetic. Encyclopedia of Animal Behavior, 55–62

Cantlon, J.F., & Brannon, E.M. (2007). Basic Math in Monkeys and College Students. PLoS Biology, vol. **5**, no. 12, December, pp. 55–62

Cantlon, J.F. (2012). Math, monkeys, and the developing brain. Proceedings of the National Academy of Sciences, vol. **109**, pp. 10725–10732

Carvalho, L.S., Knott, B., Berg, M.L., Bennett, A.T.D., & Hunt, D.M. (2010). Ultraviolet-sensitive vision in long-lived birds. Proceedings of the Royal Society B, July, pp. 1–8

Chittka, L., & Geiger, L. (1995). Can honey bees count landmarks? Animal Behaviour, vol. **49**, pp. 159–164

Davis, H., & Pérusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. Behavioral and Brain Sciences, vol. **11**, pp. 561–615

Dehaene, S. (1992). Varieties of numerical abilities. Cognition, vol. **44**, pp. 1–42

Diamond, J., & Bond, A.B. (1999). Kea, Bird of Paradox. Berkeley: University of California Press.

Doucet, M.S., Mennill, D.J., Montgomerie, R., Boag, P.T., & Ratcliffe, L.M. (2004). Achromatic plumage reflectance predicts reproductive success in male black-capped chickadees. Behavioral Ecology, vol. **16**, no. 1, September, pp. 218–222

Drake, P., Mix, K., & Clearfield, M. (2000). Precursors to Numbers: Making the Most of Continuous Amount, in Gleitman, L. R./Joshi, A. K. (eds.), Proceedings of the Twenty-Second Annual Conference of the cognitive science society. Philadelphia: Cognitive Science Society, pp. 639–643

Emmerton, J. (2001). Birds' Judgments of Number and Quantity, in: Robert G. Cook (ed.),

References

    Avian Visual Cognition, [On-line]. Available: www.pigeon.psy.tufts.edu/avc/emmerton/

Emmerton, J., & Renner, J.C. (2006). Scalar effects in the visual discrimination of numerosity by pigeons. Learning & Behavior, vol. **34**, pp. 176–192

Farnsworth, G.L., & Smolinski, J.L. (2006). Numerical discrimination by wild northern mockingbirds. The Condor, vol. **108**, pp. 953–957

Ferkin, M.H., Pierce, A.A., Sealand, R.O., & delBarco-Trillo, J. (2005). Meadow voles, *Microtus pennsylvanicus*, can distinguish more over-marks from fewer over-marks. Animal Cognition, vol. **8**, pp. 182–189

Forshaw, J.W. (2010). Parrots of the World. Princeton: Princeton University Press.

Fraser, I., & Grey, J. (2013). Australian Bird Names: A Complete Guide. Collingwood, Australia: CSIRO Publishing.

Gallistel, C.R., & Gelman, R. (2005). Mathematical cognition. In: Holyoak KJ, Morrison RG (eds.) The Cambridge handbook of thinking and reasoning. Cambridge University Press, New York, pp. 559–588

Gsell, A.C., Hagelin, J., & Brunton, D.H. (2012). Olfactory sensitivity in Kea and Kaka. Emu – Austral Ornithology, vol. **112**, no. 1, pp. 60–66

Hauser, M.D., & Spelke, E. (2004). Evolutionary and developmental foundations of human knowledge: a case study of mathematics. In: Gazzaniga M.S. (ed.) The cognitive neurosciences III. The MIT Press, Cambridge, pp. 853–864

Hirai, S., & Jitsumori, M. (2009). Counting absolute numbers of items, from 1 to 8, in pigeons. Learning & Behavior, vol. **37**, no. 4, pp. 365–379

Hunt, S., Low, J., & Burns, K.C. (2008). Adaptive numerical competency in a food-hoarding songbird. Proceedings of the Royal Society B, vol. **275**, pp. 2373–2379

Jansen, B.R., Hofman, A.D., Straatemeier, M., van Bers, B., M., C., W., Raijmakers, M.E.J., & van der Maas, H.L.J. (2014). The role of pattern recognition in children`s exact enumeration of small numbers. British Journal of Developmental Psychology, vol. **32**, pp. 187–194

Jordan, K.E., & Brannon, E.M. (2006). A common representational system governed by Weber"s law: Nonverbal numerical similarity judgments in 6-year-olds and rhesus macaques. Journal of Experimental Childhood Psychology, vol. **95**, pp. 215–229

References

Köhler, O. (1941). Vom Erlernen unbenannter Anzahlen bei Vögeln [Learning about unnamed numbers by birds]. Die Naturwissenschaften, vol. **29**, 201–218

Köhler, O. (1943). "Z hl"-versuche an einem Kolkraben und Vergleichsversuche an Menschen ["Counting"-ability in a raven and comparative research with humans]. Zeitschrift für Tierpsychologie, vol. **5**, 575–712

Köhler, O. (1951). The ability of birds to count. Bulletin of Animal Behaviour, vol. **9**, 41–45

Krajcsi, A., Szabó, E., & Mórocz, I.Á. (2013). Subitizing Is Sensitive to the Arrangement of Objects. Experimental Psychology, vol. **1**, no. 1, pp. 1–8

Krusche, P., Uller, C., & Dicke, U. (2010). Quantity discrimination in salamanders. Journal of Experimental Biology, vol. **213**, pp. 1822–1828

Mandler, G., & Shebo, B.J. (1982). Subitizing: An Analysis of Its Component Processes. Journal of Experimental Psychology. General, vol. **111**, pp. 1–22

Marriner, L.W. (2010 [1908]). The Kea: A New Zealand problem. USA: Kessinger Publishing.

McAllan, I.A.W. (2004). Corrections to the original citations and type localities of some birds described by John Gould and recorded from New Zealand. Notornis, vol. **51**, April, pp. 125–130

McIntyre, J.M., Garland, A., & Evans, T.A. (2013). What counts for „counting"? Chimpanzees, Pan troglodytes, respond appropriately to relevant and irrelevant information in a quantity judgment task. Animal Behaviour, vol. **85**, no.5, May: pp. 987–993.

Meyers, J.G. (1924). The Kea or Mountain Parrot. New Zealand Journal of Agriculture, vol. **29**

Moll, F.W., & Nieder, A. (2014). The long and the short of it: Rule-based relative length discriminationin carrion crows, *Corvus corone*. Behavoural Processes, vol. **107**: pp. 142–149

Pahl M., Si, A., & Zhang, S. (2013). Numerical cognition in bees and other insects, in Agrillo, Christian (ed.): Frontiers in Psychology, vol. **4**, April: pp. 1–9

Pepperberg, I.M. (1994). Numerical competence in an African Grey parrot. Journal of Comparative Psychology, vol. **108**, pp. 36–44

Pepperberg, I.M. (1999). The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots. Cambridge, Massachusetts, and London, England: Harvard University Press.

References

Pepperberg, I.M., & Gordon, J.D. (2005). Number Comprehension by a Grey Parrot (*Psittacus erithacus*), Including a Zero-Like Concept. Journal of Comparative Psychology, vol. **119**, No. 2, pp. 197–209

Pepperberg, I.M. (2006). Grey parrot numerical competence: a review. Animal Cognition, vol. **9**, pp. 377–391

Pepperberg, I.M. (2012). Further evidence for addition and numerical competence by a Grey parrot (*Psittacus erithacus*). Animal Cognition, vol. **15**, pp. 711–717

Petrazzini, M.E.M. (2014). Trained Quantity Abilities in Horses (*Equus caballus*): A Preliminary Investigation. Behavioral Sciences, vol. **4**, pp. 213–225

Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. Cognition (2011), doi: 10.1016/j.cognition.2011.05.007

Pulliam, H.R. (1974). On the theory of optimal diets. American Naturalist, vol. **108**, No. 959, pp. 59–74

Rahman, N.A.A., Fadzly, N., Dzakwan, N.M., & Zulkifli, N.H. (2014). The Numerical Competency of Two Bird Species (*Corvus splendens* and *Acridotheres tristis*). Tropical Life Sciences Research, vol. **25**, No.1, pp. 95–103

Range, F., Utrata, E., & Virányi, Z. (2012). Quantity discrimination in wolves (*Canis lupus*), in Beran, Michael (ed.): Frontiers in Psychology, vol. **3**, November: pp. 1–9

Revkin, S.K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? Psychological Science, vol. **19**, No.6, pp. 607–614.

Rugani, R., Regolin, L., & Vallortigara, G. (2007). Rudimental numerical competence in 5-day-old domestic chicks (*Gallus gallus*): identification of ordinal position. Journal of Experimental Psychology, vol. **33**, pp. 21–31

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, vol. **6**, pp. 147-151

Schwing, R. (2010). Scavenging behaviour of kea (*Nestor notabilis*). Notornis, vol. **57**, pp. 98–99

Steurer, M.M., Aust, U., & Huber, L. (2012). The Vienna comparative cognition technology (VCCT): An innovative operant conditioning system for various species and

experimental procedures. Behavavior Research Methods, vol. **44**, no. 4, pp. 909–918

Stevens, S.S., & Marks, L.E. ([1938] 1939). On the Problem of Scales for the Measurement of Psychological Magnitudes, in Carnap R., Reichenbach H. (ed.) The Journal of Unified Science (Erkenntnis) Volume IX. USA: University of Chicago Press.

Taves, E.H. (1941). Two mechanisms for the perception of visual numerousness. Archives of Psychology, vol. **37**, no. 265.

Toft, C.A., & Wright, T.F. (2015). Parrots of the Wild: A Natural History of the World`s Most Captivating Birds. USA, Oakland, California: University of California Press.

Ujfalussy, D.J., Miklósi, Á., Bugnyar, T., & Kotrschal, K. (2013). Role of Mental Representations in Quantity Judgments by Jackdaws (*Corvus monedula*). Journal of Comparative Psychology, September: pp. 1–10

Uller, C., Jaeger, R., Guidry, G., & Martin, C. (2003). Salamanders (*Plethodon cinereus*) go for more: rudiments of number in an amphibian. Animal Cognition, vol. **6**, pp. 105–112

Vonk, J., Torgerson-White, L., McGuire, M., Thueme, M., Thomas, J., & Beran, M.J. (2014). Quantity estimation and comparison in western lowland gorillas (*Gorilla gorilla gorilla*). Animal cognition, vol. **14**, pp. 755–765

Vonk, J. (2013). Quantity matching by an orangutan (*Pongo abelii*). Animal Cognition, vol. **6**, June: pp. 105–112

Vonk, J., & Beran, M.J. (2012). Bears "count" Too: Quantity Estimation and Comparison in Black Bears (*Ursus Americanus*). Animal Behaviour, vol. **84**, no.1, July: 231–238

Ward, C., & Smuts, B. (2007). Quantity-based judgments in the domestic dog (*Canis lupus familiaris*). Animal Cognition, vol. **10**, pp. 71–80

Watanabe, S. (1998). Discrimination of „Four" and „Two" by Pigeons. The Psychological Record, vol. **48**, pp. 383–391

Wein, A., Gajdon, G.K., & Schwing, R. (2015). Picture – Object Recognition in Kea (Nestor notabilis). Ethology, vol. **121**, pp. 1059–1070

Weser, C., & Ross, J.G. (2013). The effect of colour on bait consumption of kea (Nestor notabilis): implications for deterring birds from toxic baits. New Zealand Journal of Zoology, vol. **40**, no. 2, pp. 137-144
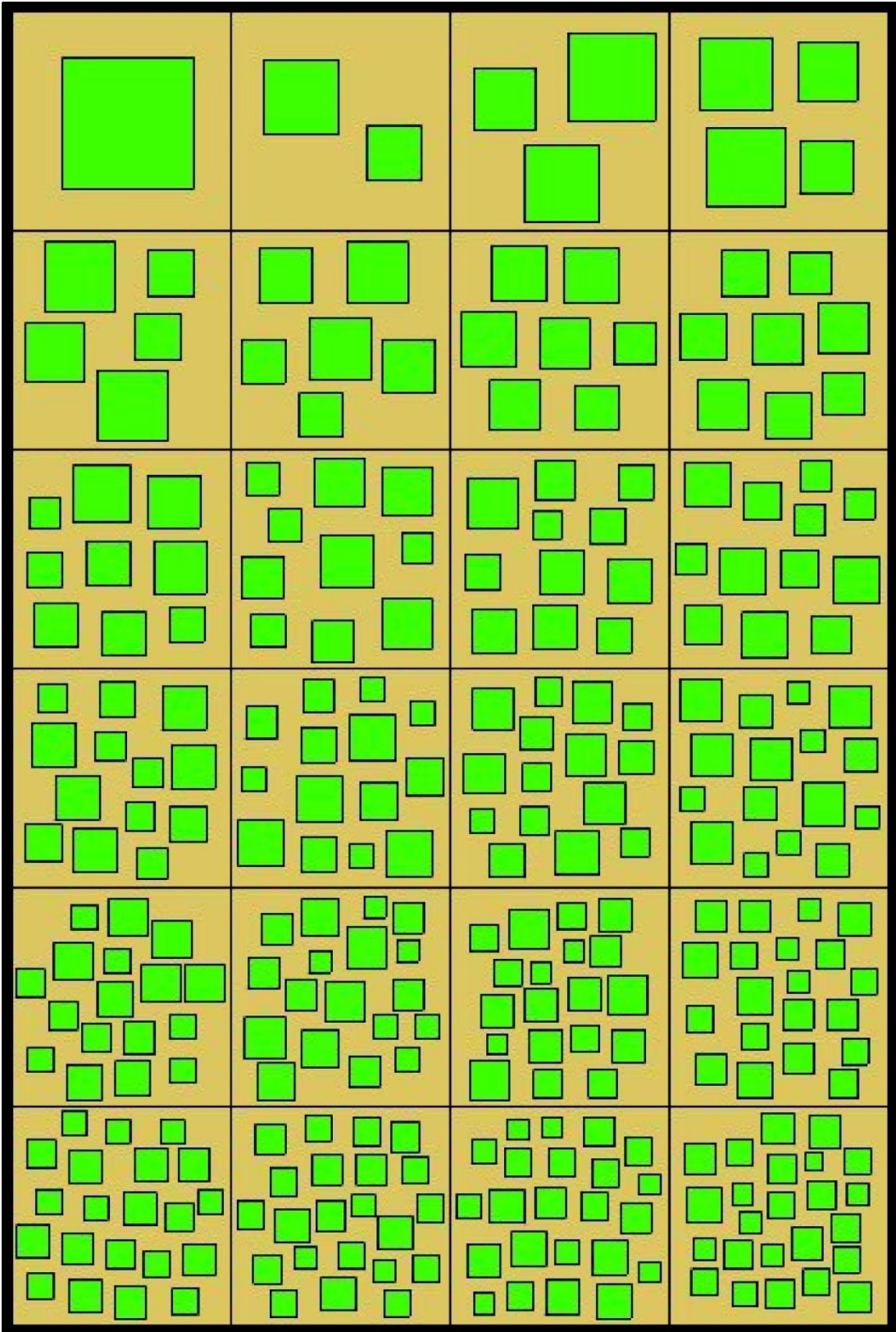
I. Appendix

I.I. Stimuli Experiment I stage II

I.I.I. All Stimuli used in Experiment I stage II version a)
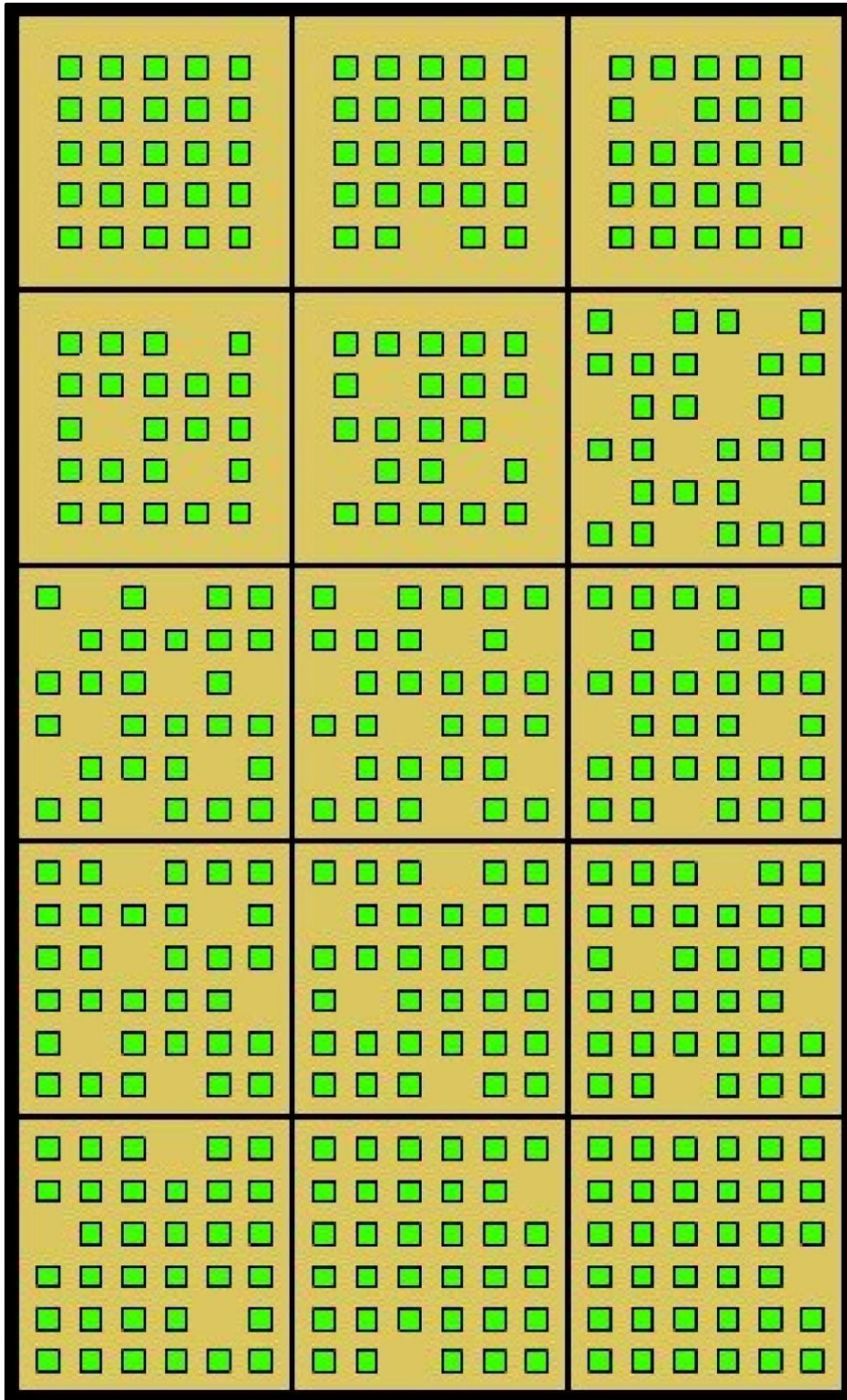
Appendix
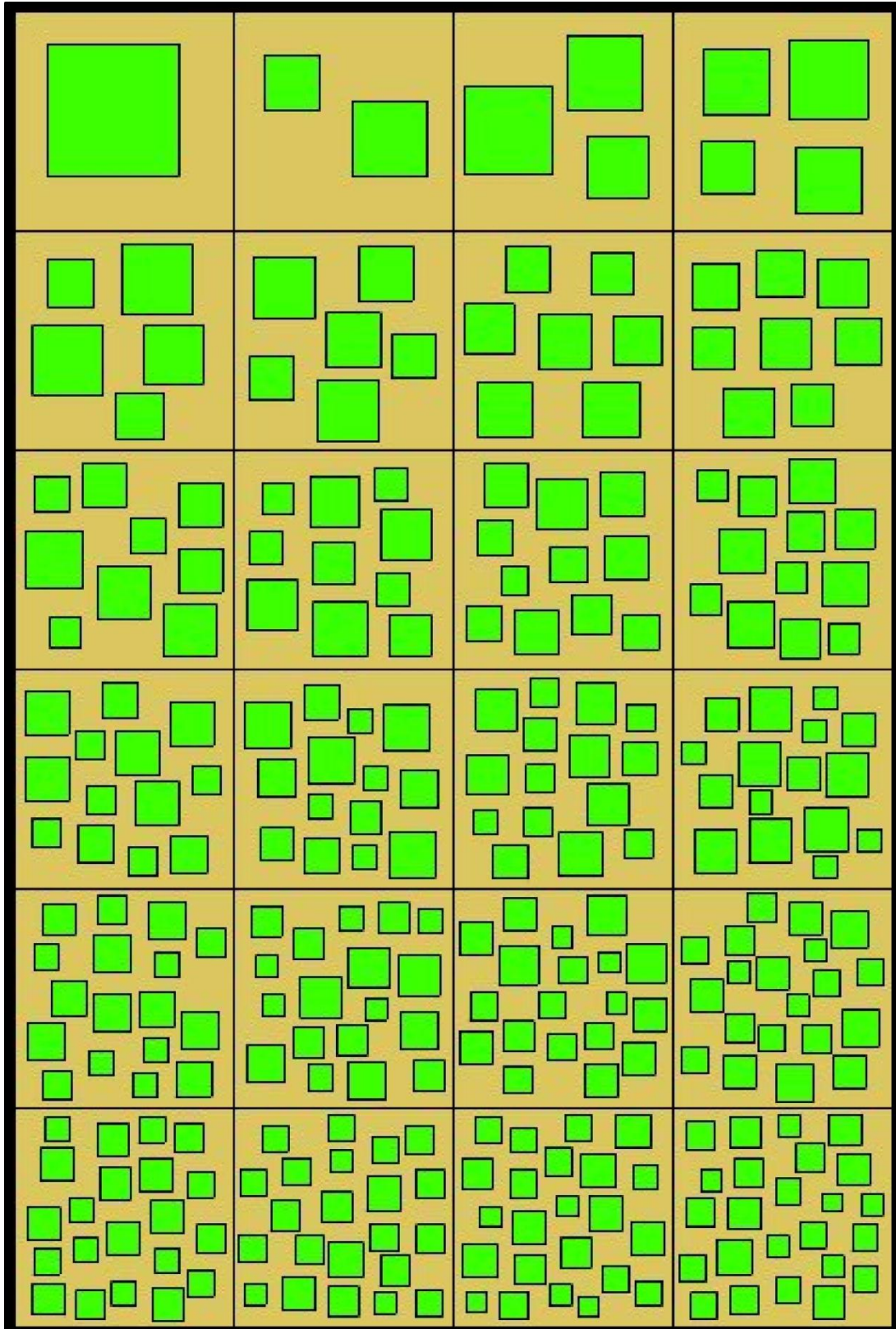
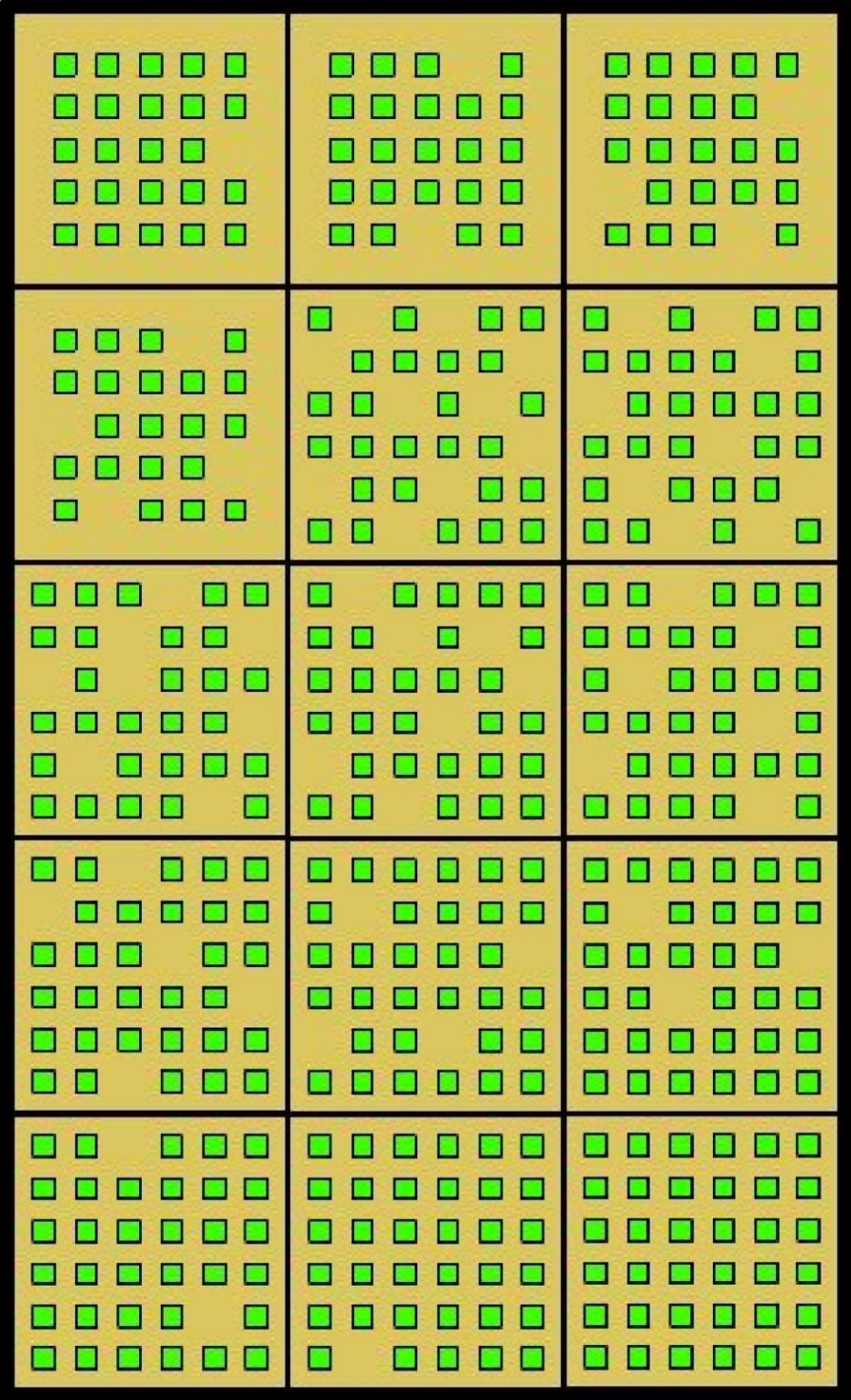**I.I.II. All Stimuli used in Experiment I stage II version b)**

I.II. Stimuli Experiment II

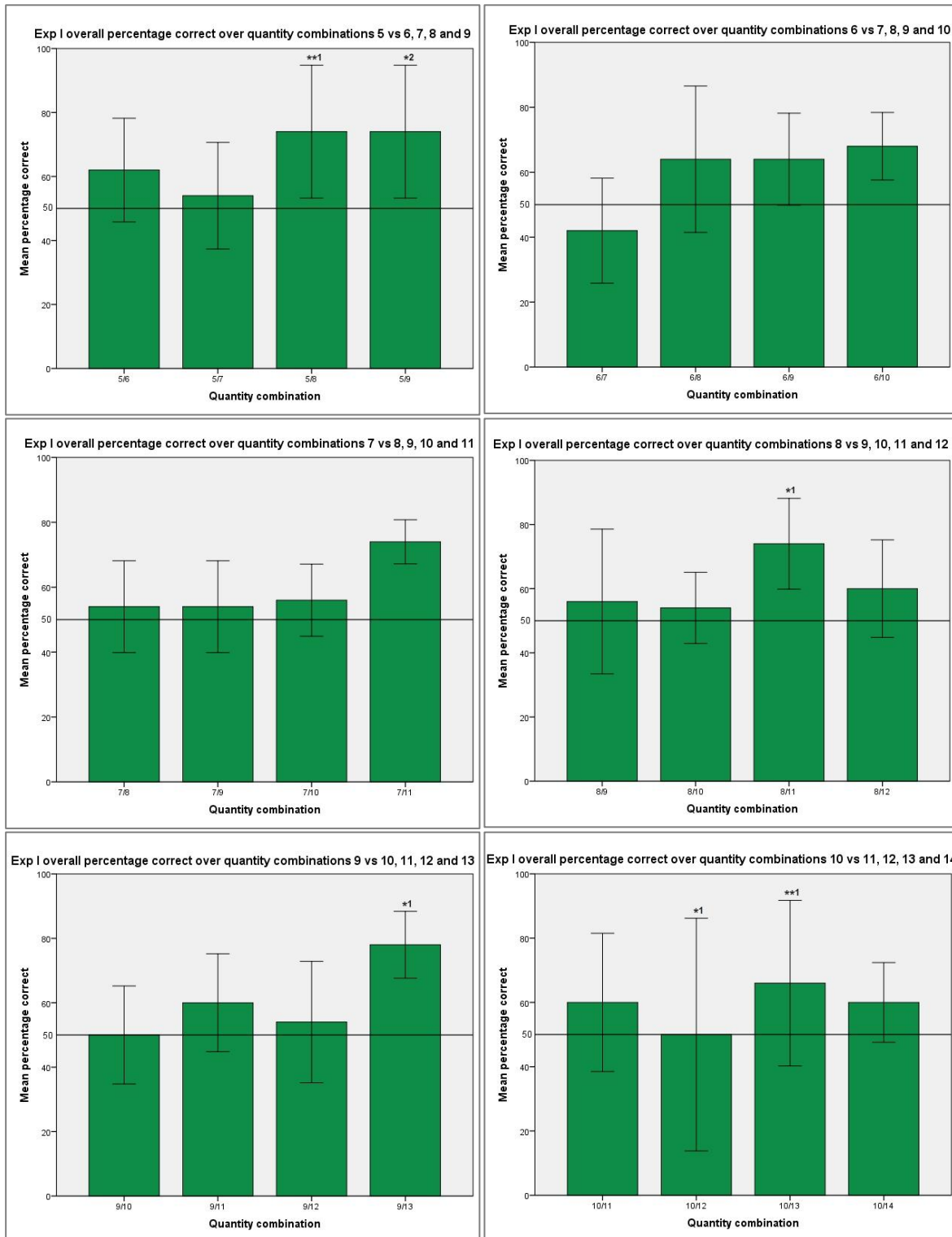### I.II.I. Stimuli used in Experiment II version a), plus quantity 25

I.II.II. Stimuli used in Experiment II version b), plus quantity 36

Appendix

I.III. Tables and Bar Graphs of Experiment I

I.III.I. Experiment I complete combinations with absolute difference from 1 to 4, the lowest quantity of the pair being constant

Appendix

61

Appendix

Exp I overall percentage correct over quantity combinations 17 vs 18, 19, 20 and 21

Exp I overall percentage correct over quantity combinations 18 vs 19, 20, 21 and 22

Exp I overall percentage correct over quantity combinations 19 vs 20, 21, 22 and 23

Exp I overall percentage correct over quantity combinations 20 vs 21, 22, 23 and 24

Appendix

I.III.II. Experiment I complete combinations with absolute differences from 1 to 4

Exp I overall percentage correct over quantity combinations with an absolute difference of 1

Exp I overall percentage correct over quantity combinations with an absolute difference of 2

Appendix

Exp I overall percentage correct over quantity combinations with an absolute difference of 3

Exp I overall percentage correct over quantity combinations with an absolute difference of 4

I.III.III. Experiment I complete combinations GLM and binomial testing

Combination	Trials	Trials Correct	B	Std. Error	Interval Lower	Interval Upper	GLM Hypothesis Test Wald Chi-Square	df	Sig.	Binomial Hypothesis Test Individual Subject(s) that reached Significance	Sig.
5 vs. 6	50	31	,490	,2914	-,082	1,061	2,823	1	,093	non-significant	> 0,05
5 vs. 7	50	27	0,16	0,2838	-0,396	0,716	0,319	1	0,572	non-significant	> 0,05
5 vs. 8	50	37	1,046	0,3224	0,414	1,678	10,525	1	0,001	Subject 3	0,001
5 vs. 9	50	37	1,046	0,3224	0,414	1,678	10,525	1	0,001	Subject 3, Subject 5	0,021
6 vs. 7	50	21	-0,323	0,2865	-0,884	0,239	1,269	1	0,26	non-significant	> 0,05
6 vs. 8	50	32	0,575	0,2946	-0,002	1,153	3,814	1	0,051	non-significant	> 0,05
6 vs. 9	50	32	0,575	0,2946	-0,002	1,153	3,814	1	0,051	non-significant	> 0,05
6 vs .10	50	34	0,754	0,3032	0,16	1,348	6,182	1	0,013	non-significant	> 0,05
7 vs. 8	50	27	0,16	0,2838	-0,396	0,716	0,319	1	0,572	non-significant	> 0,05
7 vs. 9	50	27	0,16	0,2838	-0,396	0,716	0,319	1	0,572	non-significant	> 0,05
7 vs. 10	50	28	0,241	0,2849	-0,317	0,8	0,717	1	0,397	non-significant	> 0,05
7 vs. 11	50	37	1,046	0,3224	0,414	1,678	10,525	1	0,001	non-significant	> 0,05
8 vs. 9	50	28	0,241	0,2849	-0,317	0,8	0,717	1	0,397	non-significant	> 0,05
8 vs. 10	50	27	0,16	0,2838	-0,396	0,716	0,319	1	0,572	non-significant	> 0,05
8 vs. 11	50	37	1,046	0,3224	0,414	1,678	10,525	1	0,001	Subject 5	0,021
8 vs. 12	50	30	0,405	0,2887	-0,16	0,971	1,973	1	0,16	non-significant	> 0,05
9 vs. 10	50	25	-1,89E-16	0,2828	-0,554	0,554	0,00	1	1,00	non-significant	> 0,05
9 vs. 11	50	30	0,405	0,2887	-0,16	0,971	1,973	1	0,16	non-significant	> 0,05
9 vs. 12	50	27	0,16	0,2838	-0,396	0,716	0,319	1	0,572	non-significant	> 0,05
9 vs. 13	50	39	1,266	0,3414	0,597	1,935	13,744	1	0,00	Subject 3	0,021
10 vs. 11	50	30	0,405	0,2887	-0,16	0,971	1,973	1	0,16	non-significant	> 0,05
10 vs. 12	50	25	-1,89E-16	0,2828	-0,554	0,554	0,00	1	1,00	Subject 1	0,021
10 vs. 13	50	33	0,663	0,2985	0,078	1,248	4,936	1	0,026	Subject 1	0,001
10 vs. 14	50	30	0,405	0,2887	-0,16	0,971	1,973	1	0,16	non-significant	> 0,05
11 vs. 12	50	28	0,241	0,2849	-0,317	0,8	0,717	1	0,397	non-significant	> 0,05
11 vs. 13	50	34	0,754	0,3032	0,16	1,348	6,182	1	0,013	Subject 1	0,021
11 vs. 14	50	35	0,847	0,3086	0,242	1,452	7,538	1	0,006	Subject 3	0,021
11 vs. 15	50	31	0,49	0,2914	-0,082	1,061	2,823	1	0,093	Subject 3	0,001
12 vs. 13	50	27	0,16	0,2838	-0,396	0,716	0,319	1	0,572	non-significant	> 0,05
12 vs. 14	50	26	0,08	0,2831	-0,475	0,635	0,08	1	0,777	non-significant	> 0,05
12 vs. 15	50	32	0,575	0,2946	-0,002	1,153	3,814	1	0,051	Subject 5	0,021
12 vs. 16	50	29	0,323	0,2865	-0,239	0,884	1,269	1	0,26	non-significant	> 0,05
13 vs. 14	50	26	0,08	0,2831	-0,475	0,635	0,08	1	0,777	non-significant	> 0,05
13 vs. 15	50	24	-0,08	0,2831	-0,635	0,475	0,08	1	0,777	non-significant	> 0,05
13 vs. 16	50	34	0,754	0,3032	0,16	1,348	6,182	1	0,013	non-significant	> 0,05
13 vs. 17	50	40	1,386	0,3536	0,693	2,079	15,374	1	0,00	Subject 2	0,001
14 vs. 15	50	32	0,575	0,2946	-0,002	1,153	3,814	1	0,051	non-significant	> 0,05
14 vs. 16	50	34	0,754	0,3032	0,16	1,348	6,182	1	0,013	non-significant	> 0,05
14 vs. 17	50	33	0,663	0,2985	0,078	1,248	4,936	1	0,026	non-significant	> 0,05
14 vs. 18	50	46	2,442	0,5213	1,421	3,464	21,951	1	0,00	Subjects 1,2,5 / Subject 3	0,001 / 0,021
15 vs. 16	50	31	,490	,2914	-,082	1,061	2,823	1	,093	Subject 3	0,021
15 vs. 17	50	38	1,153	0,3311	0,504	1,802	12,117	1	0,00	Subject 4	0,021
15 vs. 18	50	35	0,847	0,3086	0,242	1,452	7,538	1	0,006	non-significant	> 0,05
15 vs. 19	50	40	1,386	0,3536	0,693	2,079	15,374	1	0,00	Subject 2	0,021
16 vs. 17	50	35	0,847	0,3086	0,242	1,452	7,538	1	0,006	non-significant	> 0,05
16 vs. 18	50	24	-0,08	0,2831	-0,635	0,475	0,08	1	0,777	non-significant	> 0,05
16 vs. 19	50	39	1,266	0,3414	0,597	1,935	13,744	1	0,00	Subject 3	0,021
16 vs. 20	50	40	1,386	0,3536	0,693	2,079	15,374	1	0,00	Subject 5	0,021
17 vs. 18	50	25	-1,89E-16	0,2828	-0,554	0,554	0,00	1	1,00	non-significant	> 0,05
17 vs. 19	50	39	1,266	0,3414	0,597	1,935	13,744	1	0,00	Subject 2	0,001
17 vs. 20	50	36	0,944	0,315	0,327	1,562	8,991	1	0,003	Subject 3	0,021
17 vs. 21	50	32	0,575	0,2946	-0,002	1,153	3,814	1	0,051	non-significant	> 0,05
18 vs. 19	50	31	,490	,2914	-,082	1,061	2,823	1	,093	non-significant	> 0,05
18 vs. 20	50	24	-0,08	0,2831	-0,635	0,475	0,08	1	0,777	non-significant	> 0,05
18 vs. 21	50	35	0,847	0,3086	0,242	1,452	7,538	1	0,006	non-significant	> 0,05
18 vs. 22	50	36	0,944	0,315	0,327	1,562	8,991	1	0,003	Subject 2, Subject 5	0,021
19 vs. 20	50	23	-0,16	0,2838	-0,716	0,396	0,319	1	0,572	non-significant	> 0,05
19 vs. 21	50	27	0,16	0,2838	-0,396	0,716	0,319	1	0,572	Subject 3	0,021
19 vs. 22	50	31	,490	,2914	-,082	1,061	2,823	1	,093	non-significant	> 0,05
19 vs. 23	50	38	1,153	0,3311	0,504	1,802	12,117	1	0,00	Subject 3, Subject 4	0,021
20 vs. 21	50	38	1,153	0,3311	0,504	1,802	12,117	1	0,00	Subject 2	0,021
20 vs. 22	50	36	0,944	0,315	0,327	1,562	8,991	1	0,003	Subject 2	0,001
20 vs. 23	50	44	1,992	0,4352	1,139	2,845	20,96	1	0,00	Subject 2 / Subjects 4, 5	0,001 / 0,021
20 vs. 24	50	40	1,386	0,3536	0,693	2,079	15,374	1	0,00	Subject 2 / Subject 1	0,001 / 0,021

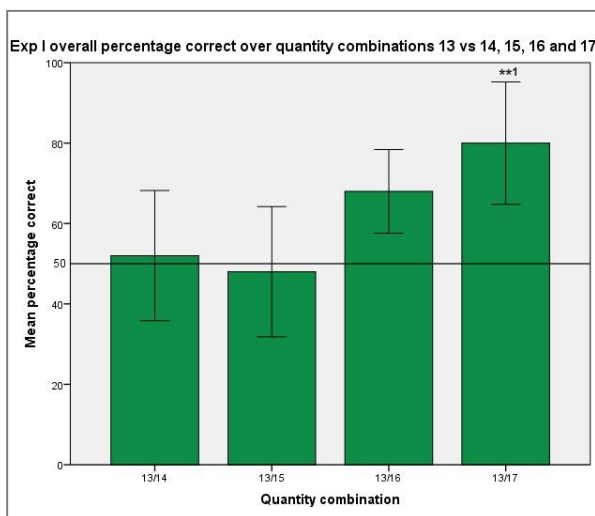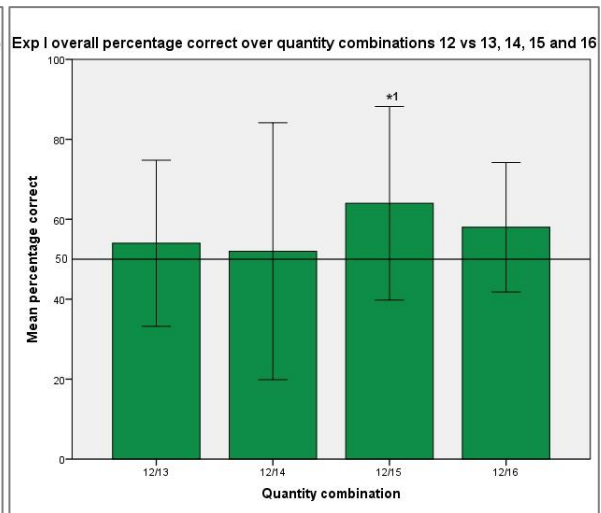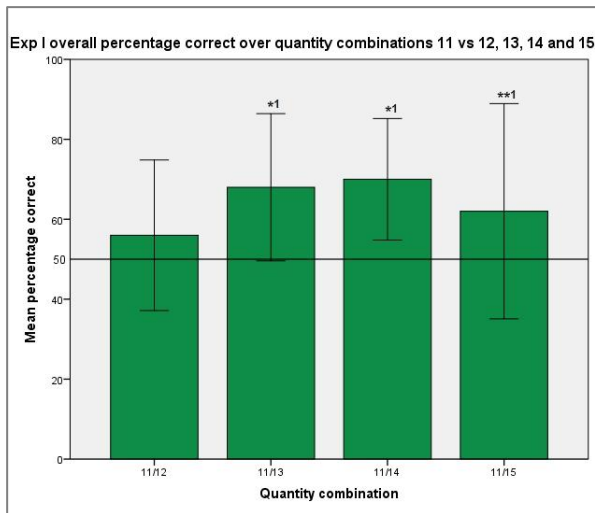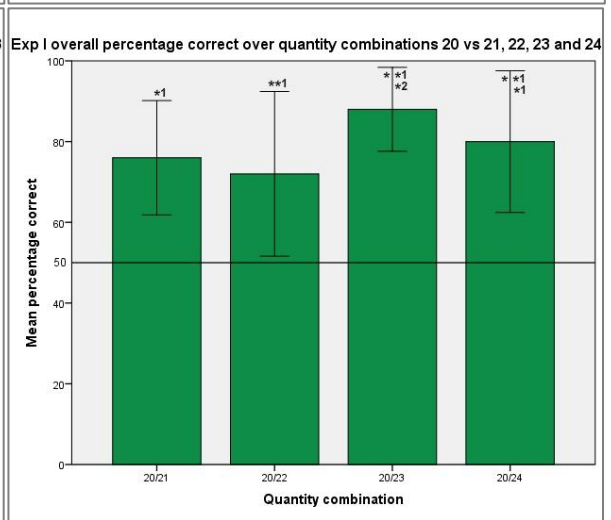## I.IV. Tables and Bar Graphs of Experiment I stage II
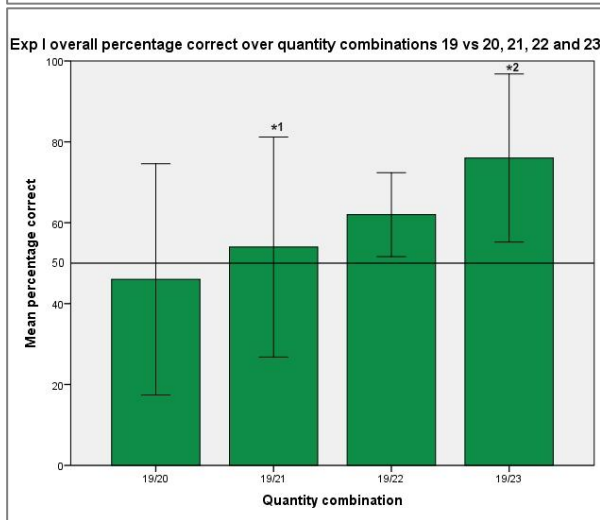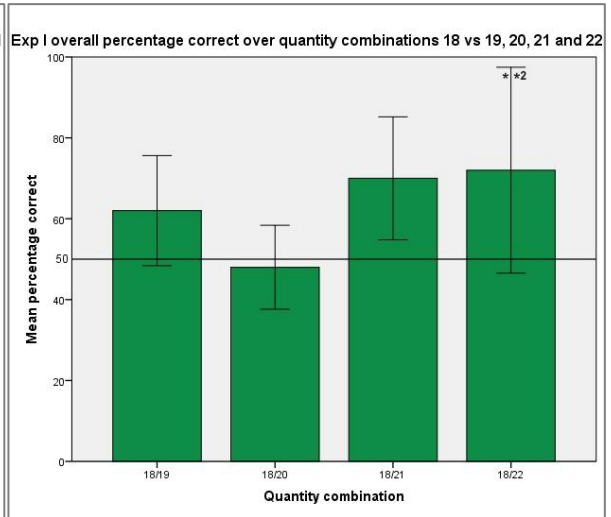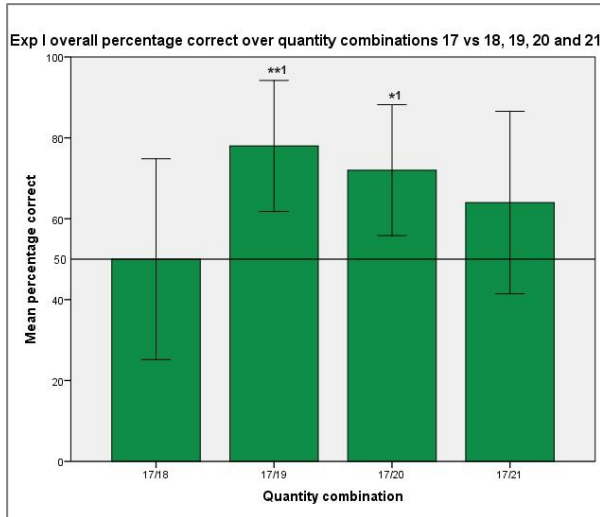
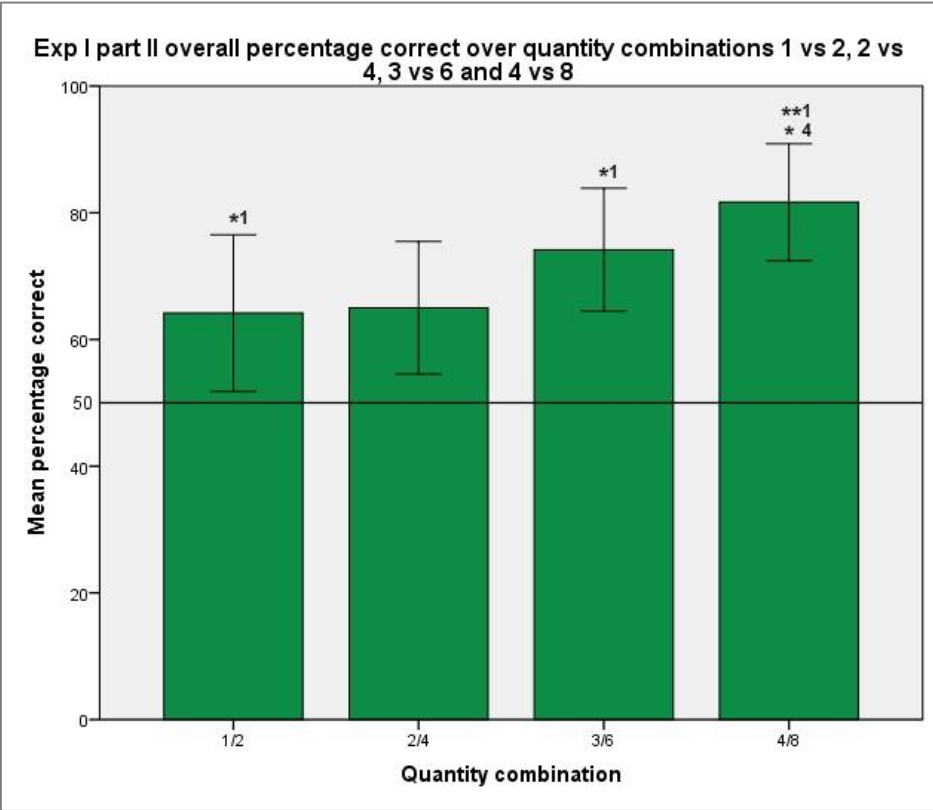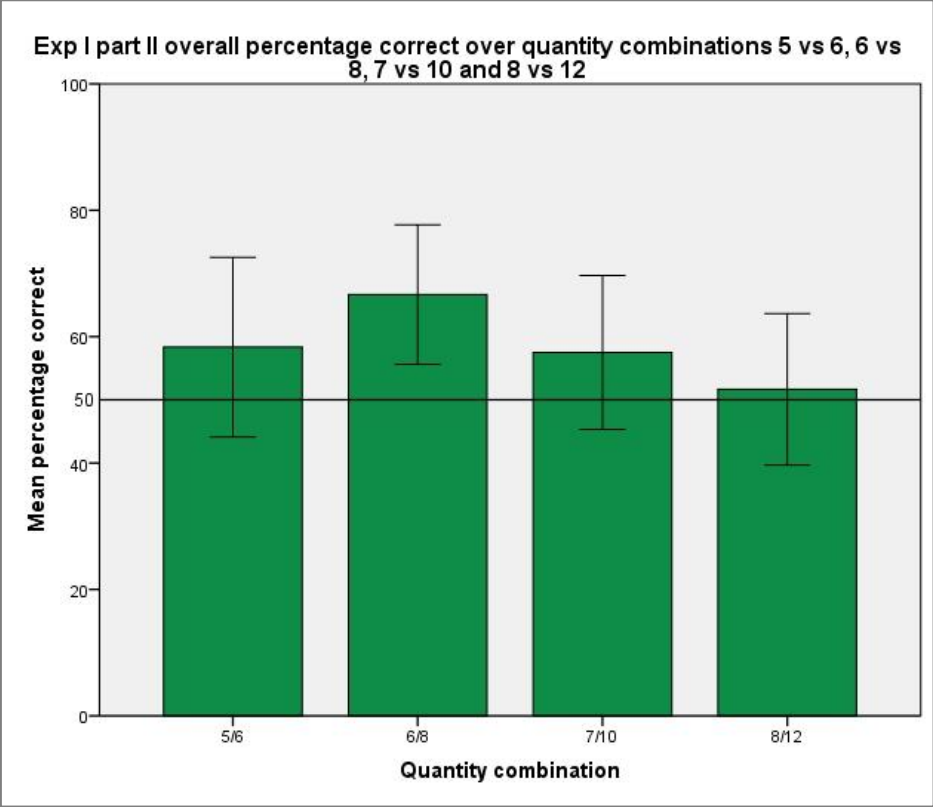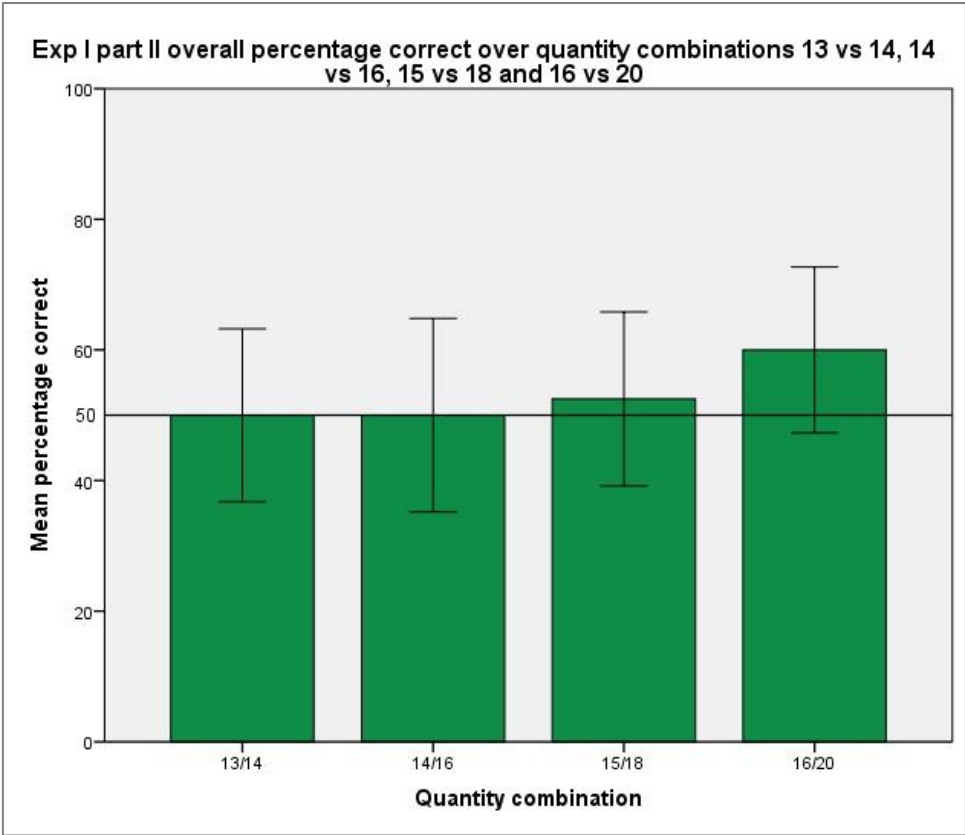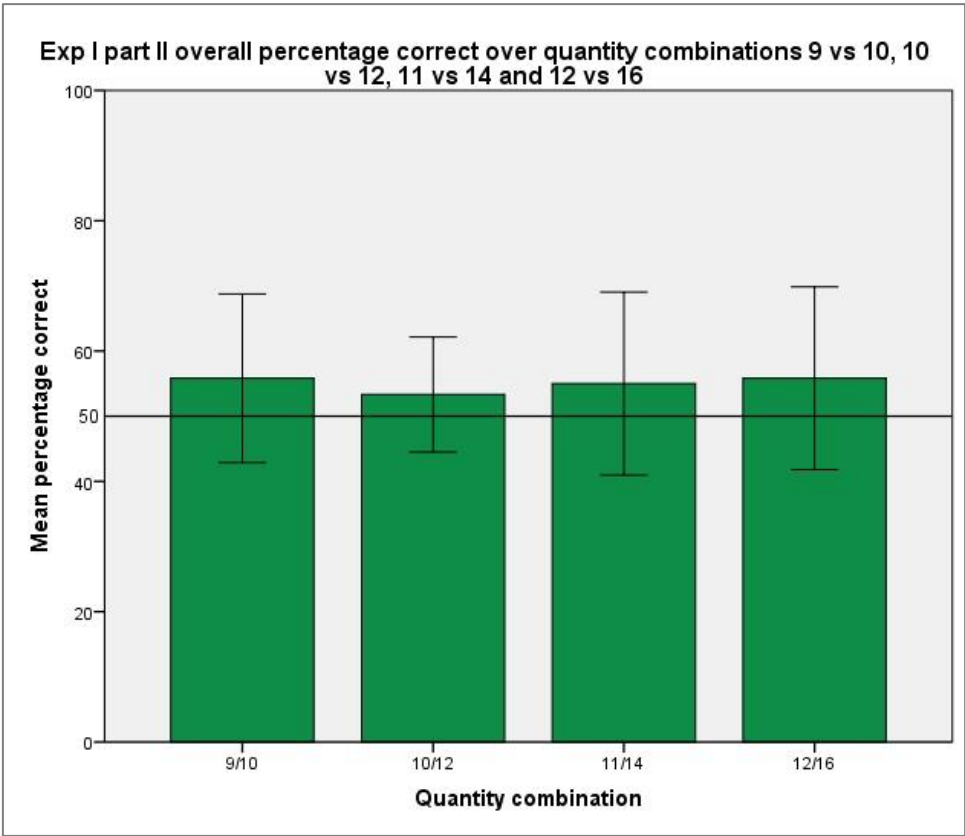### I.IV.I. Experiment I stage II complete combinations with absolute differences ranging from 1 to 4

Exp I part II overall percentage correct over quantity combinations 9 vs 10, 10 vs 12, 11 vs 14 and 12 vs 16

Exp I part II overall percentage correct over quantity combinations 13 vs 14, 14 vs 16, 15 vs 18 and 16 vs 20

I.IV.II. Experiment I stage II complete combinations with absolute differences 1 to 4

Exp I stage II overall percentage correct over quantity combinations with an absolute difference of 1

Exp I stage II overall percentage correct over quantity combinations with an absolute difference of 2

Exp I stage II overall percentage correct over quantity combinations with an absolute difference of 3

Exp I stage II overall percentage correct over quantity combinations with an absolute difference of 4

69

I.IV.III. Experiment I stage II complete combinations GLM and binomial testing

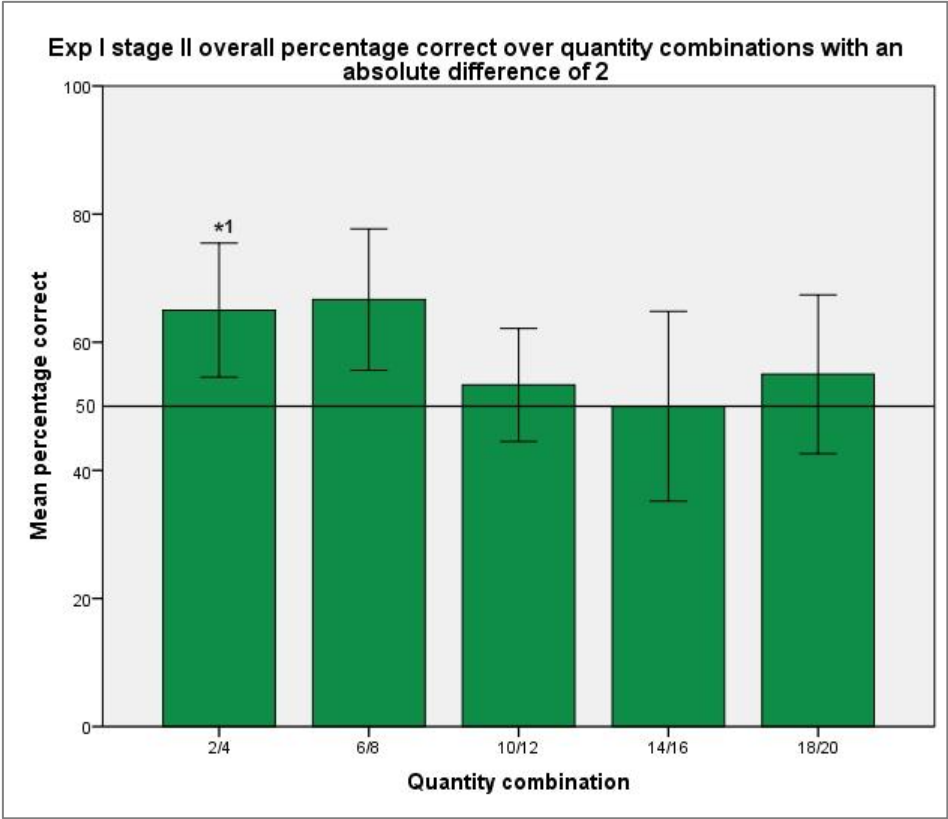| Combination	Trials	Trials Correct	B	Std. Error	Interval		Wald Chi-Square	df	Sig.	Individual Subject(s) that reached Significance	Sig.
					Lower	Upper					
1 vs. 2	120	77	0,583	0,1904	0,209	0,956	9,365	1	0,002	Subject 10	0,021
2 vs. 4	120	78	0,619	0,1914	0,244	0,994	10,462	1	0,001	Subject 3	0,021
3 vs. 6	120	89	1,055	0,2086	0,646	1,463	25,573	1	0,00	Subject 9	0,021
4 vs. 8	120	98	1,494	0,2359	1,032	1,956	40,098	1	0,00	Subject 12 / Subjects 5, 9, 10, 13	0,001 / 0,021
5 vs. 6	120	70	0,336	0,1852	-0,026	0,699	3,302	1	0,069	non-significant	> 0,05
6 vs. 8	120	80	0,693	0,1936	0,314	1,073	12,812	1	0,00	non-significant	> 0,05
7 vs. 10	120	69	0,302	0,1847	-0,06	0,664	2,68	1	0,102	non-significant	> 0,05
8 vs. 12	120	62	0,067	0,1827	-0,291	0,425	0,133	1	0,715	non-significant	> 0,05
9 vs. 10	120	67	0,234	0,1838	-0,126	0,595	1,626	1	0,202	Subject 13	0,021
10 vs. 12	120	64	0,134	0,183	-0,225	0,492	0,533	1	0,466	non-significant	> 0,05
11 vs. 14	120	66	0,201	0,1835	-0,159	0,56	1,196	1	0,274	non-significant	> 0,05
12 vs. 16	120	67	0,234	0,1838	-0,126	0,595	1,626	1	0,202	non-significant	> 0,05
13 vs. 14	120	60	7,36E-18	0,1826	-0,358	0,358	0,00	1	1,00	non-significant	> 0,05
14 vs. 16	120	60	-3,80E-17	0,1826	-0,358	0,358	0,00	1,00	1,00	non-significant	> 0,05
15 vs. 18	120	63	0,10	0,1828	-0,258	0,458	0,30	1	0,584	non-significant	> 0,05
16 vs. 20	120	72	0,405	0,1863	0,04	0,771	4,735	1	0,03	non-significant	> 0,05
17 vs. 18	120	54	-0,201	0,1835	-0,56	0,159	1,196	1	0,274	non-significant	> 0,05
18 vs. 20	120	66	0,201	0,1835	-0,159	0,56	1,196	1	0,274	non-significant	> 0,05
19 vs. 22	120	52	-0,268	0,1842	-0,629	0,093	2,121	1	0,145	non-significant	> 0,05
20 vs. 24	120	63	0,10	0,1828	-0,258	0,458	0,30	1	0,584	non-significant	> 0,05

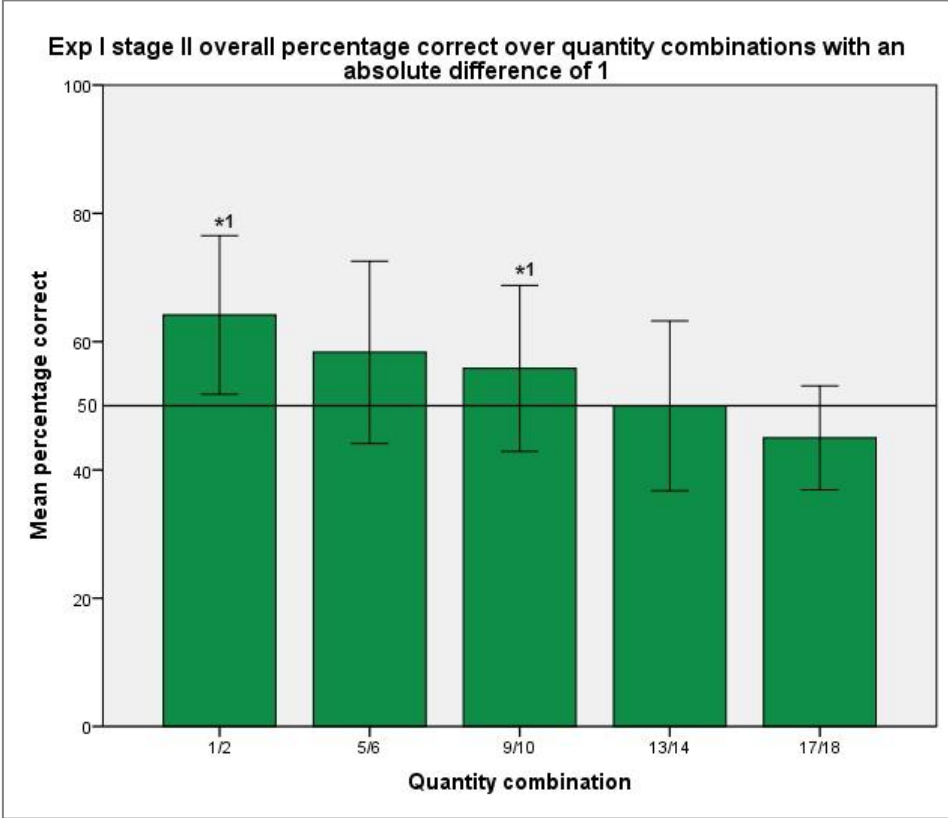# I.V. Tables and Bar Graphs of Experiment II

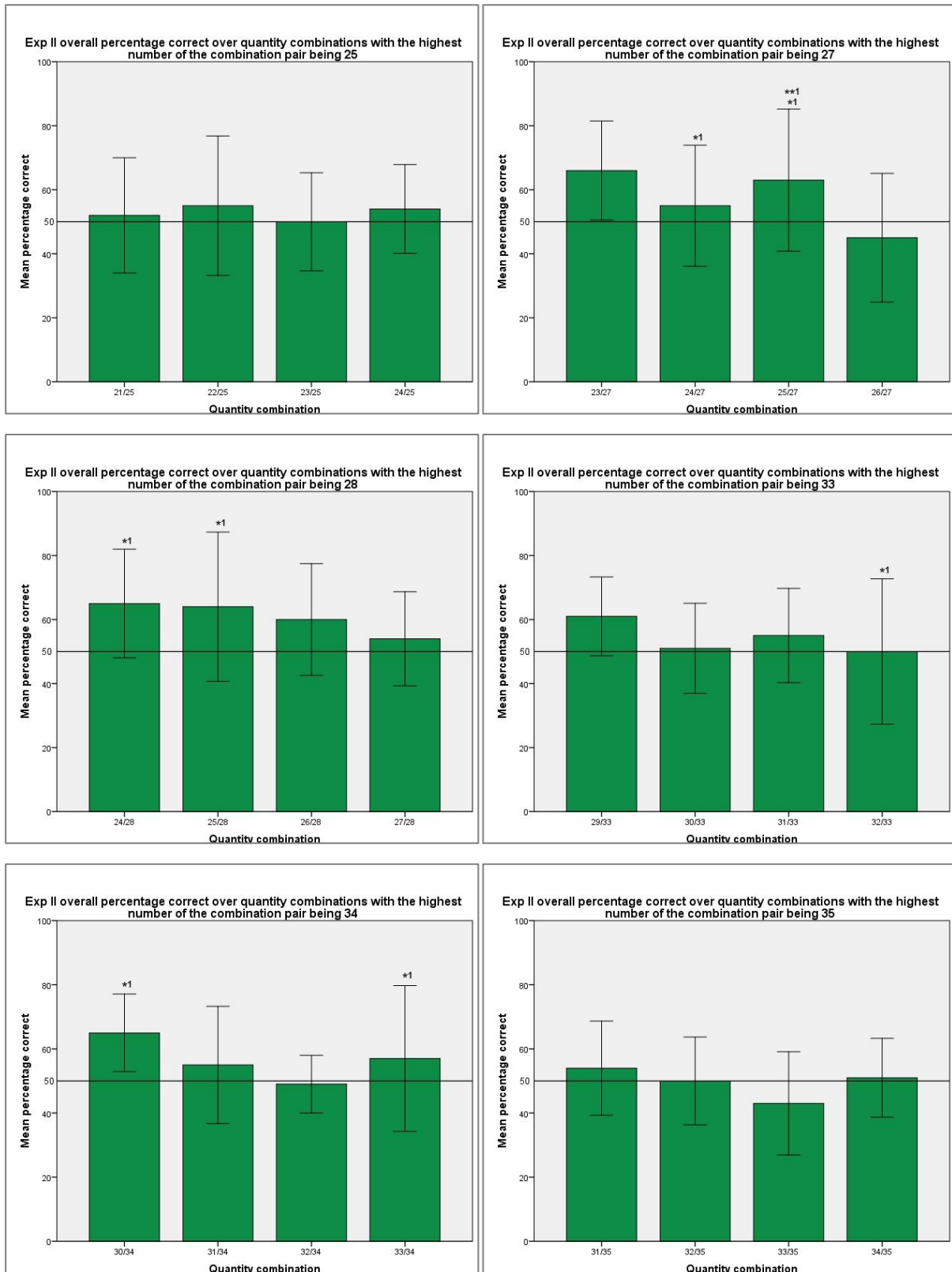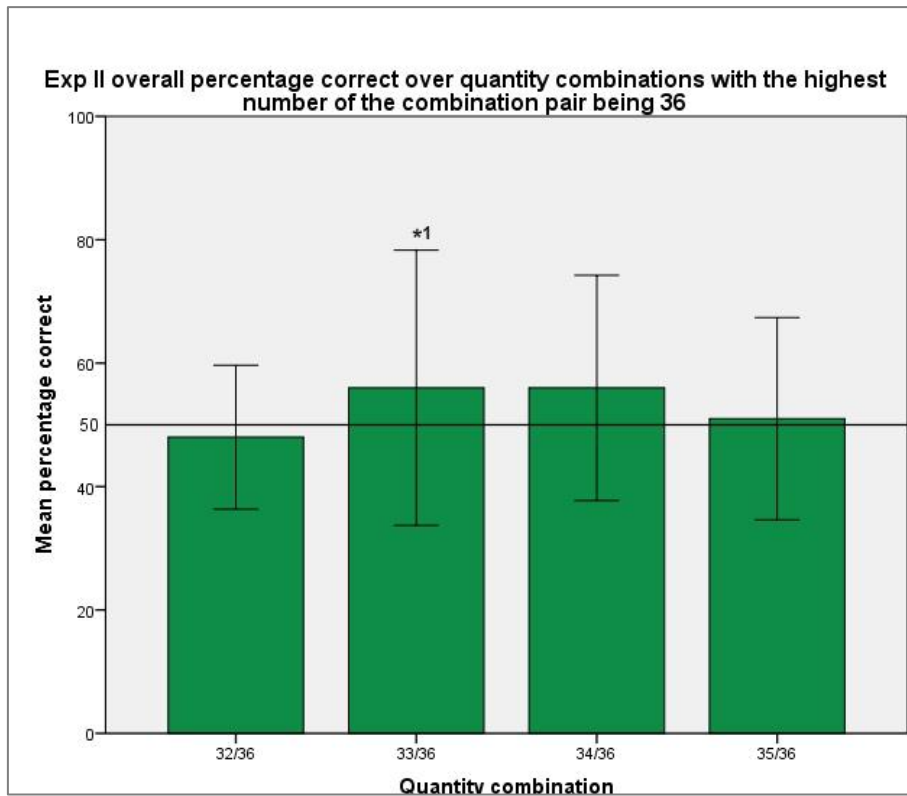## I.V.I. Experiment II complete combinations with absolute difference from 1 to 4, the highest quantity of the pair being constant

Exp II overall percentage correct over quantity combinations with the highest number of the combination pair being 36

I.V.II. Experiment II complete combinations with absolute differences from 1 to 4

Exp II overall percentage correct over quantity combinations with an absolute difference of 1

Exp II overall percentage correct over quantity combinations with an absolute difference of 2

Exp II overall percentage correct over quantity combinations with an absolute difference of 3
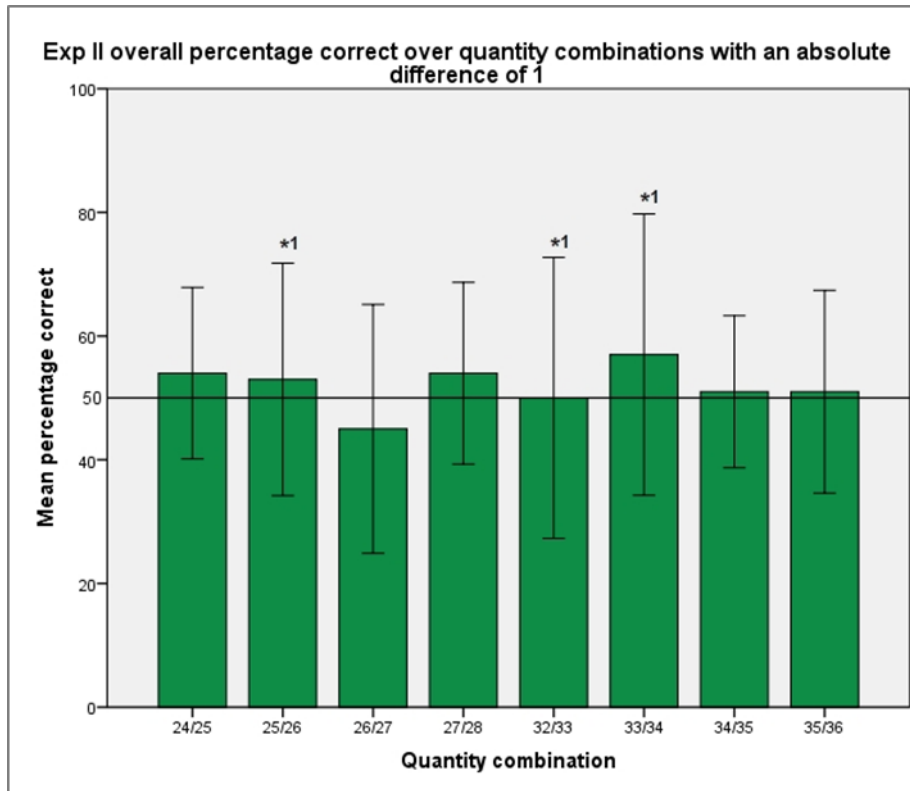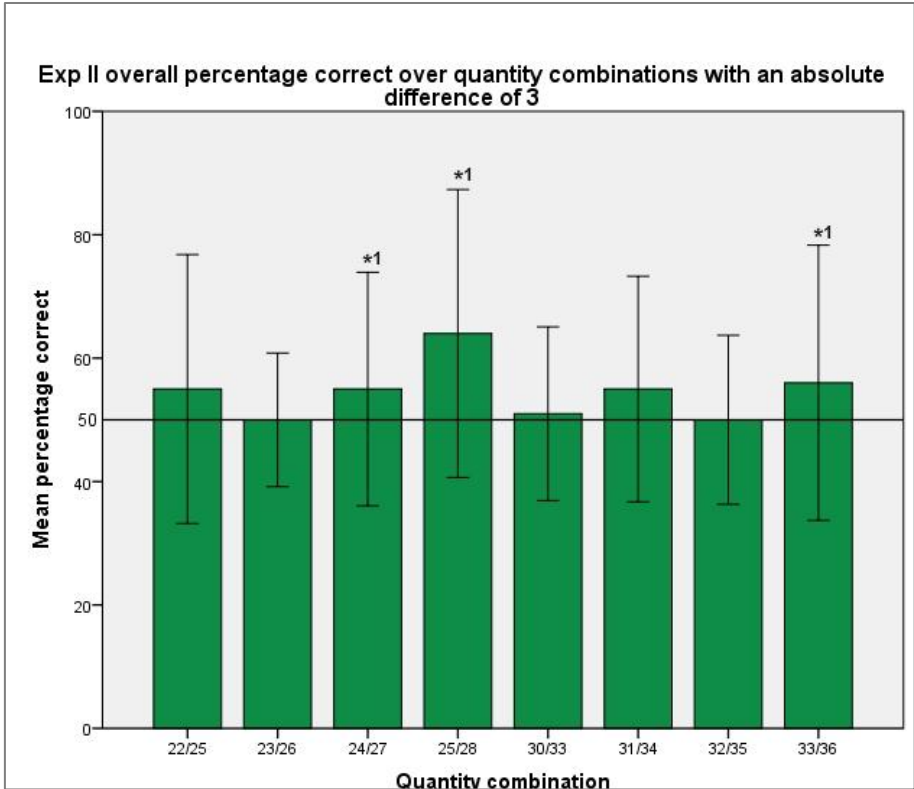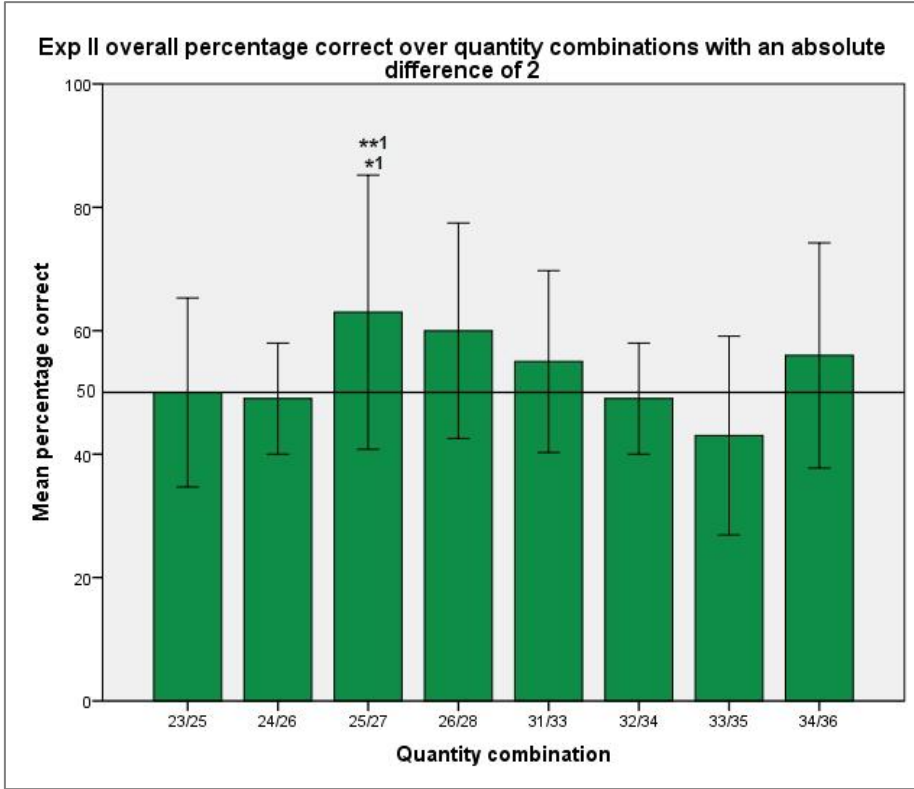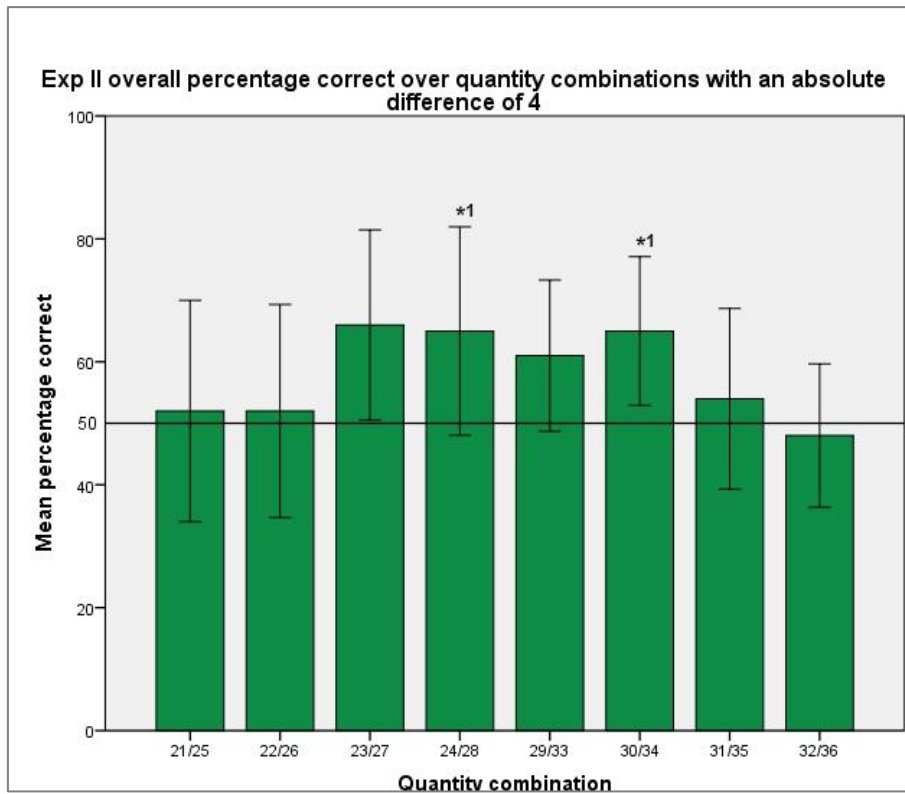
I.V.III. Experiment II complete combinations GLM and binomial testing

Combination	Trials	Trials Correct	B	Std. Error	95% Wald Confidence Interval		GLM Hypothesis Test			Binomial Hypothesis Test	
					Lower	Upper	Wald Chi-Square	df	Sig.	Individual Subject(s) that reached Significance	Sig.
21 vs. 25	110	60	.182	.1915	-.193	.558	.907	1	.341	non-significant	> 0,05
22 vs. 25	110	64	.330	.1933	-.049	.709	2.919	1	.088	non-significant	> 0,05
22 vs. 26	110	58	.109	.1910	-.265	.484	.327	1	.567	non-significant	> 0,05
23 vs. 25	110	55	-3,11E-14	.1907	-.374	.374	.000	1	1.000	non-significant	> 0,05
23 vs. 26	110	55	-3,78E-14	.1907	-.374	.374	.000	1	1.000	non-significant	> 0,05
23 vs. 27	110	70	.560	.1982	.171	.948	7.972	1	.005	non-significant	> 0,05
24 vs. 25	110	59	.146	.1912	-.229	.520	.581	1	.446	non-significant	> 0,05
24 vs. 26	110	55	-1,59E-13	.1907	-.374	.374	.000	1	1.000	non-significant	> 0,05
24 vs. 27	110	58	.109	.1910	-.265	.484	.327	1	.567	Subject 2	0,021
24 vs. 28	110	72	.639	.2005	.246	1.032	10.159	1	.001	Subjekt 6	0,021
25 vs. 26	110	58	.109	.1910	-.265	.484	.327	1	.567	Subject 2	0,021
25 vs. 27	110	67	.443	.1954	.061	.826	5.151	1	.023	Subject 2 / Subject 10	0,001 / 0,021
25 vs. 28	110	71	.599	.1993	.208	.990	9.036	1	.003	Subjekt 6	0,021
26 vs. 27	110	54	-.036	.1907	-.410	.337	.036	1	.849	non-significant	> 0,05
26 vs. 28	110	62	.256	.1923	-.121	.633	1.772	1	.183	non-significant	> 0,05
27 vs. 28	110	58	.109	.1910	-.265	.484	.327	1	.567	non-significant	> 0,05
29 vs. 33	110	64	.330	.1933	-.049	.709	2.919	1	.088	non-significant	> 0,05
30 vs. 33	110	59	.146	.1912	-.229	.520	.581	1	.446	non-significant	> 0,05
30 vs. 34	110	70	.560	.1982	.171	.948	7.972	1	.005	Subject 2	0,021
31 vs. 33	110	63	.293	.1927	-.085	.671	2.311	1	.128	non-significant	> 0,05
31 vs. 34	110	60	.182	.1915	-.193	.558	.907	1	.341	non-significant	> 0,05
31 vs. 35	110	59	.146	.1912	-.229	.520	.581	1	.446	non-significant	> 0,05
32 vs. 33	110	54	-.036	.1907	-.410	.337	.036	1	.849	Subject 2	0,021
32 vs. 34	110	54	-.036	.1907	-.410	.337	.036	1	.849	non-significant	> 0,05
32 vs. 35	110	54	-.036	.1907	-.410	.337	.036	1	.849	non-significant	> 0,05
32 vs. 36	110	53	-.073	.1908	-.447	.301	.145	1	.703	non-significant	> 0,05
33 vs. 34	110	61	.219	.1918	-.157	.595	1.304	1	.254	Subjekt 6	0,021
33 vs. 35	110	49	-.219	.1918	-.595	.157	1.304	1	.254	non-significant	> 0,05
33 vs. 36	110	59	.146	.1912	-.229	.520	.581	1	.446	Subjekt 8	0,021
34 vs. 35	110	57	.073	.1908	-.301	.447	.145	1	.703	non-significant	> 0,05
34 vs. 36	110	61	.219	.1918	-.157	.595	1.304	1	.254	non-significant	> 0,05
35 vs. 36	110	57	.073	.1908	-.301	.447	.145	1	.703	non-significant	> 0,05

I.VI. Curriculum Vitae

David Zdravko Beraha

Born: 06.08.1987

Nationality: Austria

Place of residence: Vienna

E-Mail: david.beraha@outlook.com

Education

- **Rupert Riedl Price for Master`s Thesis in Social and Cultural Anthropology (2016)**

Messerli Research Institute
(Vienna, Austria)

(University of Vienna, Medical University of Vienna, Veterinary University of Vienna)

- **Master of Science - "Interdisciplinary Master in Human-Animal interactions" (WS 2013–)**

Subjects include:

Animal Ethics, Animal Welfare, Comparative Cognition, Comparative Medicine, Law

University of Vienna _____(Vienna, Austria)

- **Master of Arts - "Social and Cultural Anthropology" (SS 2013 – WS 2016) Graduation with Distinction**

Subjects include:
Medical Anthropology, Philosophy of Science, History of Science, Philosophical Anthropology, Anthropology of Media, Anthropology of Politics, Ethnographic Practice, Historical Anthropology, Anthropology of Knowledge, Power and Ethics, Anthropology and Human Rights, Anthropology of Modernity, Regional Cultures of India, Visual Anthropology, Anthropology of Sound and Language

Regional foci: India, Africa, Oceania

- **Bachelor of Arts – "Social and Cultural Anthropology" (WS 2010 – WS 2013)**

Subjects include:
Ethnomusicology, Visual Anthropology, Medical Anthropology, Anthropology of Religion, Anthropology of Consciousness, Anthropology of Nature, Anthropology of Law, Historical Anthropology, Anthropology of Gender, Colonialism, Racism and Ethnicity, Kinship studies, Urban Anthropology, Neuro-Anthropology, Empirical Methods (Interview, Statistics, Fieldwork), Ethics

Regional foci: New Zealand/Aotearoa, South-East Asia, Mesoamerica

Work Experience

- **IT Help Desk at Caritas Österreich**

  (September 1st, 2019 - , Vienna, Austria)

  - IT Service Desk, Data Management, Technical support

- **IT Support at Business University Vienna (WU-Wien)**

  (April 10th - July 1st, 2019, Vienna, Austria)

  - IT Support, Equipment service, Technical support, Event Management

- **Vienna Anthropology Days (VANDA)**

  (September 19th - 22nd, 2018, Vienna, Austria)

  - Corporate Identity, Press and Marketing, Social Media Administration

- **Assistent in financial accounting at Südwind – Verein für Entwicklungspolitik und globale Gerechtigkeit**

  (April 1st, 2017 - January 1st, 2018, Vienna, Austria)

  - Assistance in finances, book-keeping, databank administration, accounting and billing

- **Civil servant at Südwind – Verein für Entwicklungspolitik und globale Gerechtigkeit**

  (July 1st, 2016 - March 31st, 2017, Vienna, Austria)

  - Assistance in various areas such as bureau management, finances, information services, communication, public relations, courier services, databank administration

- **Eleventh Conference on Hunting and Gathering Societies (CHAGS 11)**

  (September 7th-11th, 2015, Vienna, Austria)

  Assistance in various areas such as material preparation, information desk and room attendance