Variation in response behaviours in captive common marmosets (Callithrix jacchus)

Nicholas Simon Blackwood

Variation in Response Behaviours in Captive Common Marmosets (*Callithrix jacchus*)

Nicholas Simon Blackwood

2006

Evolutionary Anthropology Research Group
Department of Anthropology
43 Old Elvet
Durham
DH1 3HN

Submitted to the University of Durham for the degree of Doctor of Philosophy

Nicholas Simon Blackwood

Variation in Response Behaviours in Captive Common Marmosets (*Callithrix jacchus*)

Abstract

Individual variation can be seen in many aspects of an organism, from its physical structure to its behaviour. Contributing factors to variation in behaviour may include sex, age, genetic differences and even size. The response to new objects and environments is a varying behavioural trait found in a wide range of species. The aim of this thesis was to investigate the causes of variation in response to novel stimuli in the common marmoset, *Callithrix jacchus*. The investigation focused on the effects of sex, age, genetic differences and size. Variation in response was tested by a simple novel stimulus presentation test paradigm. Sixty eight animals were each individually presented with nine novel stimuli in home cage tests. Five measures of response were recorded: latency to approach and contact, duration of proximity and contact, and visual attendance. Responses were analysed and stimuli were categorised as: mirror, food related stimuli, unattractive stimuli and novel stimuli. Response across the nine stimuli was investigated for variation due to sex, age or weight of the subjects. Across the analysis, limited significant sex differences were seen in response to food related stimuli, with males being more responsive. To investigate whether general measures of response could be derived from the individual behaviours recorded, principal component analysis was carried out on the response data, which was split into the four stimulus groups. Simple response continua were successfully derived from components from analysis of mean stimulus group scores. The mirror and food stimulus groups each had two continua, one reflecting latency to response, and one reflecting the duration of time spent near the stimulus. The responses to the unattractive stimulus group and novel stimulus group could each be described by one response continuum. In order to assess whether genetic variation contributed to response, heritability analyses were carried out on both the derived continua and the five response measures, separated by stimulus group. No significant heritabilities were found after correction for multiple comparisons. This study thus demonstrates that sex is a more important determinant of response that individual genetic differences, age or weight in the common marmoset.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost I thank my supervisor Jan De Ruiter, for allowing me to work with him on this project, and for his ideas and enthusiasm. I am deeply indebted to Russell Hill, not only for his help with statistical analyses and his inability to refuse to proofread chapters for me, but also for supplying me with lodgings and beer whenever they were required. Thank you to Hilary Box, Geoff Hosey and Todd C Rae for advice during my viva and suggestions for my corrections. I would also like to thank my co-supervisor Robert Barton, and the Evolutionary Anthropology Research group for many stimulating coffee time discussions about nothing to do with evolutionary anthropology.

I would like to thank Peter Pearce and Leah Scott of Dstl for their input into the project, and the provision of a study group. I would not have been able to carry out my work with the marmosets if it were not for the help and advice of David Veall, Sean Hearson and the husbandry staff in the breeding colony. Thanks also to all of the Porton scientists for allowing me to have far too many tea breaks with them, and to Dan Stevens for supplying me with lifts, heavy metal and bits of information that I forgot to take away with me.

My time at Durham would not have been what it was but of the companionship, help and advice of my office mates Kerrie Lewis, Louise Blundell, Trudi Buck, Bryony Whiting, and other friends (some now long gone) including Sarah, Adam, Becca, Rob, Erik, Vasco, Lucy, Andy, Ian and the mathmos, and probably some other people who I have forgotten.

Thank you to my family for supporting me, not only financially (at times), but also in the not telling me to "stop being a student and get a real job" sense. And finally, I would like to thank Beccy, not only for moving all the way to Durham to put up with the grim Northern weather and constant conversations about monkeys, but also for being the finest contents list editor I have ever had the pleasure to work with.

No one supposes that all the individuals of the same species are cast in the very same mould.
(Darwin, 1859: 102)

Chapter One

Introduction

1.1 The causes of variation in behaviour

Individual variation can be seen in many aspects of an organism, from its physical structure to its behaviour. Individuals vary in how they respond to their environment, and this variation in behaviour can be caused by many factors. Contributing factors to variation in behaviour may include sex, age, genetic differences and even size. Social factors, such as dominance relationships, or kinships of members of a group of animals can also affect behaviour (Joubert & Vauclair, 1986; Drea, 1998). The aim of this thesis is to investigate the causes of variation in response to novel stimuli in the common marmoset, *Callithrix jacchus*. The investigation focuses on the effects of sex, age, genetic differences and size, attempting to avoid social influences by testing animals individually.

1.1.1 Sex

Physical and behavioural variation due to the sex of an animal can affect all aspects of its life. For instance, in fiddler crabs (*Uca tangeri*), where males have one greatly enlarged claw, sex differences exist in evasive behaviour (Jordão & Oliveira, 2001). Bird species (such as zebra finches, *Taeniopygia guttata*), show sexual dimorphism in song structure, which is based on underlying physical systems (Wade & Arnold; 2004). Psychological assessments of hyenas (*Crocuta crocuta*) show that males are more highly strung than females (Gosling & John, 1999). In primates, there are many differences in both behaviour and physiology related directly to sex (Dixson, 1998), and other marked behavioural differences influenced less directly by it. Male play, for instance, is more vigorous and frequent than female play (Fagen, 1981). Chimpanzee (*Pan troglodytes*) males have been rated as more 'aggressive' and 'gregarious' than females, while females were more 'timid' and 'controlled' than males (Buirski *et al.*, 1978; c.f. Smuts, 1987). Female vervet monkeys (*Cercopithecus aethiops*) are more 'opportunistic' than males (McGuire *et al.*, 1994). Higher risk-taking behaviour in

males has been discussed with reference to emigration from the natal group during young adulthood in rhesus macaques (*Macaca mulatta*) (Mehlman *et al.*, 1995).

Such variation, coupled with differences in life history between males and females, can lead to predictions of differences in specific behaviours. As males have higher variation in their reproductive success compared to females (known as Bateman's principle) it can be expected that males will in general take greater risks (Bateman 1948; Futuyma, 1998). In terms of responsiveness, for instance, this would suggest that males would be more inclined to investigate novel objects or situations.

1.1.2 Age

Patterns in life history can constrain or promote the behaviour of an animal at a particular age. Play, for example, begins in infancy, increases in juvenile animals and declines steadily during adolescence and into adulthood; immature primates spend a great deal of non-feeding time engaged in social play (Walters, 1987). The social relations of animals change as they grow, ultimately relating to the sociosexual system of the species (Walters, 1987). In many animal species, one or both sexes disperse at some point during their life, because of competition for food or mates, and to avoid inbreeding (Hewitt & Butlin, 1997). With dispersal comes a suite of behavioural changes. When male rhesus macaques reach an age at which they disperse, they become more aggressive, less sociable, and more likely to take risks (Mehlman *et al.*, 1995). These indirect behavioural correlates show individual variation; less aggressive, more sociable animals will stay in their natal group for longer. In fact, this individual variation could be a consistent cause of behavioural differences as important as age or sex.

1.1.3 Individual variation

Individual variation in behaviour is reported in many cases to be both consistent and measurable. Rhesus macaques can be consistently described as 'uptight' or 'laidback' (Suomi, 1991) and 'possessive' versus 'relaxed' maternal styles in macaques run in families (Maestriperi *et al.*, 1997). Human (*Homo sapiens*) children at either end of the normal spectrum of behaviours can be consistently labelled as 'bold' or 'shy' (Kagan *et al.*, 1988; Kagan & Snidman, 1991).When examined, some of these consistent,

measurable differences in behaviours (and characteristics based on behaviours) can be shown to have a genetic basis.

In The Descent of Man, Darwin (1871: 110) states:

> in regard to mental qualities, their transmission is manifest in our dogs, horses and other domestic animals. Besides special tastes and habits, general intelligence, courage, bad and good tempers, etc. are certainly transmitted.

Studies in human populations have examined the genetics and heritability of the psychological trait of novelty seeking (Benjamin *et al.*, 1996; Ebstein *et al.*, 1996), which is based (in part) on an inclination towards exciting behaviours (Hamer & Copeland, 2000). These studies, involving a combination of interviews, personality questionnaires and genetic tests of blood samples, show that higher than average scores in novelty seeking are found in individuals that have a certain allele in the locus for the dopamine receptor gene D4R. Furthermore, these differences are due to genetic transmission rather than population stratification. The variation found, however, only explains about 10% of the genetic variation seen in novelty seeking (Hamer & Copeland, 1998). Variation in the same region has also been found in other great apes and gibbons (Inoue-Murayama *et al.*, 2000; Shimada *et al.*, 2004).

Behaviours, and the neurochemicals associated with them, have also been shown to have a genetic influence in non-human primates. Behaviours reflecting increased stress responses and behavioural inhibition have found to be heritable in rhesus macaques (Williamson *et al.*, 2003). Genetic variation has been indicated in response to a 'social' stimulus (i.e., a mirror) in baboons (*Papio hamdryas*) (Rogers *et al.*, 2002). Several monoamine metabolites related to the neurotransmitters serotonin, dopamine and noradrenaline, which are related to personality and individual variation in psychological traits in humans, are also heritable in baboons (Rogers *et al.*, 2004). Variation in response to social stimulus (an intruder challenge test) in vervet monkeys is correlated with 5-HIAA, a serotonin metabolite (lower levels are related to an increase in impulsive behaviours), and there is a genetic component to vervet social impulsivity and aggressiveness (Fairbanks *et al.*, 2004). Thus, social impulsivity is both heritable and related to neurotransmitter levels.

The studies cited above use response behaviours to test the heritability of psychological traits. As a genetic influence on responsiveness is seen in humans (Benjamin *et al.*, 1996; Ebstein *et al.*, 1996), non-human primates (Fairbanks *et al.*, 2004), and other mammals, for instance mice (*Mus musculus*) (Flint *et al.*, 1995; Duluwa *et al.*, 1999), it could be hypothesised that all primate species would show heritable variation in response behaviour.

1.1.4 Physical variation and behaviour

As well as demographic factors such as age and sex affecting behaviour, it has been suggested that variation could reflect different behavioural strategies (Wilson *et al.*, 1994). Animals of a different physical type may use different strategies; for instance, large animals may behave qualitatively differently to small animals. This occurs in extreme form with relation to male sexual strategies in many species (e.g., orang-utan, *Pongo pygmaeus*: Utami *et al.*, 2002; coho salmon, *Oncorhynchus kisutch*: Gross, 1985; ruff, *Philogamus pugnax*: van Rhijin, 1973, 1983). Such variation has been suggested in a less extreme form for 'boldness' as a behavioural trait (Wilson *et al.*, 1994). In its simplest form, variation can take the form of a shy-bold continuum across a population. In a more complicated system, there could be innately and fixedly 'bold' or 'shy' individuals at either end of the behavioural continuum, but most of the population being phenotypically plastic and less extreme (Wilson *et* al., 1994, Wilson & Yoshimura, 1994). If such behavioural strategies are linked to physical variation, as with male sexual strategies, a measure such as weight may co-vary with responsiveness scores. For instance, in three-spined sticklebacks (*Gasterosteus aculeatus*), 'bold' individuals grow more over the same time period than 'shy' individuals (Ward *et al.*, 2004), which leads to bold adult sticklebacks being heavier than shy adults. The occurrence of such strategies is not necessarily mutually exclusive with individual variation, as it may be that the evolutionary advantage of an array of responses is what drives the variation among conspecifics. If there are simple explanatory variables such as an animal's weight, however, they should be accounted for and explained.

1.2 Responsiveness and response behaviours

Many of the behaviours discussed above are related to responsiveness. To investigate the causes of variation in response behaviours, it is necessary to define exactly what is

meant by the term, and how best it can be studied. All animals respond to novelty in some way.

Bardo *et al.* (1996, p25) state that:

> (o)rganisms are biologically prepared to attend to novel information more readily than familiar information. At the sensory level, the visual, auditory, olfactory and tactile systems are designed such that stimuli lose their impact with constant of repetitive presentation. This process of sensory adaptation or habituation biases the system towards reacting to novel stimuli, as well as to increases in the intensity of a familiar stimulus

Response to novelty has been explored in many ways, in a great variety of animal species. The human psychological trait of novelty or thrill seeking involves "the desire to seek out new experiences or thrills" (Hamer & Copeland, 1998: 11) and "exhilaration or excitement in response to novel stimuli" (Cloninger, 1987: 574). Individuals who are highly novelty seeking would thus respond more positively to novel stimuli or environments than individuals who are less novelty seeking. Similarly, many animal species show intraspecific variation in 'shyness' and 'boldness' (Wilson *et al.*, 1994).

1.2.1 Defining response behaviours

There are many definitions of behaviour related to response and responsiveness, often depending on the theoretical field in which the study is based. For example, 'emotionality' in mice has been investigated by looking at responses to standard tests that examine response to novel and stressful situations (Flint *et al.*, 1995). Emotionality in this case was assessed by movement and faecal dropping rate. A whole suite of animal species have been assessed as 'bold' or 'shy' (Wilson *et al.*, 1994). The temperament of an animal can be assessed by observing its behaviours and responses to others (Clarke & Boinski, 1995). Perhaps to avoid the subjective influence of terminology, many authors simply use the word 'response' (e.g. Box, 1988; Fairbanks, 2001; Fragaszy & Mason, 1978; Hardie & Buchanan-Smith, 2000; Visalberghi *et al*; 2003). Response, however, is not a simple thing. As Hardie & Buchanan-Smith (2000: 630) state:

> Responses to novelty consist of complex mixtures of factors that vary according to the species involved. They may include attributes of the object-

> colour, shape, size, patterning - the animal's lifestyle and prior experience, their cognitive abilities and social grouping.

To limit the aspects of response studied, the term neophilia has been used to describe an animal's attraction to a new object or task (Day *et al.*, 2002), contrasted with neophobia, a specific avoidance of new stimuli or environments (Corey, 1978). Neophilia and neophobia are not necessarily part of one continuum, but could instead be two conflicting behaviours (Greenberg, 2003). The terms can be defined (Greenberg, 2003: 179) as:

> Neophilia is the spontaneous attraction of an animal to a food object, object or place because it is novel...Neophobia is the aversion that an animal displays towards approaching a food item, object, or place simply because it is novel.

Due to these strict definitions neophilia and neophobia could be seen as unambiguous descriptors of response when novelty is involved. Unfortunately, if they represent two separate behavioural mechanisms working against each other to cause a balanced response (Greenberg 2003), the clarity is lost. An animal that fails to approach a new stimulus could be either neophobic or merely lacking in neophilia. Equally, if an animal is much quicker than its group-mates in approaching a new food source, is it neophilia in play, or is a deficiency in neophobia putting the animal at risk of poisoning? As a result, response is again left as the best, or least problematic, descriptor.

If responsive is a positive reaction, the definition of 'unresponsive' is problematic. In this discussion, and throughout this thesis, unless specifically stated otherwise, response is taken to be positive, explorative behaviour, and less responsive or unresponsive behaviour is interpreted as a lack of reaction to a stimulus or situation, as opposed to a negative (i.e., stressful; or fearful) response.

1.3 Studying response behaviours in primates

Several different areas of research overlap in their ability to suggest what may affect primate response behaviours. Behavioural and genetic investigations have demonstrated variation in response to social stimuli (Fairbanks, 2001; Fairbanks *et al.*, 2004; Rogers *et al.*, 2002). Studies concerned with animal welfare have shown that novelty plays a significant role in the use of play structures (Taylor *et al.*, 1997), and

that the addition of novel objects can increase activity in some singly housed individuals (Line & Morgan, 1991).

Other studies have specifically looked at response to novelty. For example, after seven days of presentations to captive Guinea baboons (*Papio hamadryas papio*), discovery times for novel objects stabilised, possibly indicating habituation to the method of object presentation (Joubert & Vauclair, 1986). No differences in discovery times were seen between 'natural' and 'artificial' objects. Mean frequencies of exploratory and manipulatory responses showed significant differences between females and juveniles for new objects and both new and repeatedly presented objects together. Juveniles consistently demonstrated higher levels of exploratory behaviour (staring and sniffing).

Dominance position and other social factors may affect reaction to novel objects. This is something that must be taken into account when novel objects are introduced to individuals in a group situation. The dominant male in a baboon group, for example, participated least in novelty related activities (Joubert & Vauclair, 1986). Social context also affects how rhesus macaques explore their environment (Drea, 1998). When a captive rhesus group is split into dominant and submissive, the submissive group are less likely to explore a novel food containing stimulus than when the group is together. Dominant individuals, however show no difference in their reaction. Factors other than dominance can also affect response to novel stimuli. When titi monkeys (*Callicebus moloch*) in female/male pairs are tested together and individually, both sexes are distressed in the absence of their cage-mate, and interest in novel objects is reduced (Fragaszy & Mason, 1978). Squirrel monkeys (*Saimiri sciureus*), however, show no distress when separated, and females increase their response to novel objects. Other species differences are also apparent; generally, squirrel monkeys are "quicker and more vigorous" than titis in their behaviour (Fragaszy & Mason, 1978: 311) in similar situations.

New World primates, especially capuchins (*Cebus*), have been the focus of research into the characteristics and function of response to novelty in feeding (Fragaszy *et al.*, 1997; Visalberghi & Fragaszy, 1995; Visalberghi *et al.*, 2003). Wild white-fronted capuchins (*C. apella*) are less responsive toward generic novel objects than toward novel foods (Visalberghi *et al.*, 2003) and animals respond more slowly to experimental

presentations than to known foods. Males in the wild are more persistent in their interactions with novel foods than females are (Visalberghi *et al.*, 2003), but females are more behaviourally active towards "potential foraging substrate" (i.e., more likely to come into contact with something that might contain food). The authors suggest that capuchins display a combination of explorative and neophobic behaviours that result in a gradual exploitation of new resources (this is echoed by Greenberg, 2003, in his discussion of the interplay between neophilia and neophobia in birds). Younger capuchins contact artificial feeding platforms first and are more responsive to novel foods (Visalberghi *et al.*, 2003). Infants pick up foods more often when they are novel, and eat novel foods more frequently than familiar ones (Fragaszy *et al.* 1997). Older infants are more likely than younger ones to pay attention to another individual's food.

1.3.1 Response in Callitrichids

Response behaviours have been studied in some depth in various callitrichid species. Callitrichids are small, diurnal, arboreal platyrrhine primates, including marmosets, tamarins and lion tamarins (Hershkovitz, 1977). Studies have taken place both in the field and in captive situations. Common marmosets are an ideal species to study in captivity because of their abundance in laboratories, their small size allowing colonies to hold large numbers of animals in good environmental conditions.

1.3.2 Species ecology and differences in responsiveness within the callitrichids

The natural environment and social system of callitrichids may affect response behaviours. Interspecific variation in response to novelty can be addressed in terms of the different ecological niches occupied by species, and also by sociosexual differences. For instance, variation in feeding ecology in Weid's marmosets (*Callithrix kuhli*) and golden lion tamarins (*Leontopithecus rosalia*) relates to variation in successful response to a memory task (Platt *et al.*, 1996). Weid's marmoset, as in most marmoset species, has a specialised anterior dentition that allows them to gouge trees to obtain gum, and an enlarged caecum to aid gum digestion (Rylands, 1989; Ferrari & Martins, 1992). Golden lion tamarins do not use gum as extensively as marmosets, but otherwise have a large overlap in food types (Rylands, 1989). In experimental memory tasks, Weid's marmosets perform significantly better than golden lion tamarins for short retention periods, but worse when the retention period is 24 hours. Because its specialised

dentition gives it better access to gum, Weid's marmosets require different temporal visuospatial abilities related to foraging; golden lion tamarins are more reliant on remembering where food patches are than are marmosets. Marmosets and tamarins show differences in response to novel objects in home cage tests over time, with common marmosets showing an increase in latency and cotton-top tamarins (*Saguinus oedipus)* a decrease (Millar *et al.,* 1988). As the authors (Millar *et al.,* 1988: 83) put it, "(m)armosets appear to become bored and tamarins less fearful with repeated presentations". This difference is attributed to differences in feeding ecology, with the more insectivorous tamarins observing stimuli for longer before approaching. In single and mixed species groups of saddleback (*Saguinus fuscicollis*) and red bellied tamarins (*Saguinus labiatus*), the latter, who normally forage higher in the canopy, respond to objects towards the top of the enclosure more quickly. Overall, saddleback tamarins respond more quickly; this can be linked to their extractive foraging style. Reaction times decrease for both species in mixed species groups (Hardie & Buchanan-Smith, 2000).

Differences also exist between the two tamarin genera (*Saguinus* and *Leontopithecus*) in response to a social stimulus, namely conspecifics of both sexes (French & Inglett, 1991). In a captive experiment, male cotton top tamarins responded agonistically to and attacked male intruders (but not females), whereas females did not differentiate between the sex of intruders and displayed "benign indifference" (French & Inglett, 1991: 283). In golden lion tamarins, females reacted quickly and aggressively to female intruders, with a high level of threat displays. These differences between species in response to an intruder are interpreted as being due to differences in sociosexual behaviour, and different mechanisms of reproductive suppression. Cotton top tamarin breeding females suppress ovulatory activity in other females; golden lion tamarins do not (French, 1987; French *et al.*, 1984). This means that a cotton-top tamarin female entering a group will not reproduce, but instead may even help with the rearing of the alpha female's offspring.

Contrasts can be seen, then, between *Callithrix*, *Leontopithecus* and *Saguinus* in their response to novelty, because of ecological and reproductive differences. It is possible that intraspecific differences may have an ecological or sociosexual basis, being caused by different behavioural strategies (Wilson *et al.*, 1994). An evolved variation in responsiveness could be stable in a population, due to frequency dependant selection, if

there is variation in habitat quality. It is possible that both flexible and inflexible response behaviour strategies could exist in a population, with specialist ultra-responsive and non-responsive individuals at either end of a behavioural continuum having facultatively flexible phenotypes in between (Wilson *et al.*, 1994; Wilson & Yoshimura, 1994).

1.3.3 Sex and variability in callitrichid responsiveness

Other fundamental aspects of an animal may have an affect on how it responds to novel stimuli and situations. Sex can be an important factor, due to reproductive strategy differences between males and females. The age of an animal and the stage of life history reached may also have consequences.

Response to food and foraging tasks

Many studies of sex differences in responsiveness in callitrichids focus on response to novel food, or foraging related tasks (Box, 2003). For example, in the majority of cases of male/female pairs of two species of tamarin (red bellied and cotton-top), females make the initial approach to a task, and females approached novel, food-reward, tasks more than males (Box 1988). In a subsequent study, Box *et al.* (1995) again found sex differences in these species, and in a third, the saddle-back tamarin, where females attempted food tasks more frequently and more successfully than males. As both males and females were present during the food tasks, males may have been deferring to the females, allowing them primary access to food. Saddleback tamarins approached tasks less frequently and for less time than the other two species.

In normal and enriched (additional, novel food) feeding times, marmoset (common marmoset and black tufted-eared marmoset, *Callithrix penicillata)* females are more responsive than males to normal food when additional food is given (Box & Smith, 1995). Females also become more responsive to the additional food over time. Female marmosets have also been shown to be more behaviourally responsive to supplemental foods, spending more time feeding on them than males (Petto & Devin, 1988), another example of priority of access to food for females in callitrichid species (Box *et al.*, 1997). Such dominance over food supplies are likely to be at its peak during lactation, when energetic demands on the female are at their highest (Nievergelt & Martin, 1999).

There are differences, however, in vision between males and some female callitrichids; all males are dichromatic and females can be either dichromatic or trichromatic (Caine & Mundy, 2000). The differences in foraging ability between males and females could be due to differences in vision, rather than sex differences *per se* (Box, 2003).

Response to novelty

Research on responsiveness in callitrichids has focussed on feeding and foraging, but some data on non food-related variation are available. When interacting with a novel environment and/or novel stimuli, few sex differences have been seen for common marmosets, for example, no sex differences are seen in common marmosets exploring both a novel environment, and a novel object in a social (home cage) situation (Rogers, 1999). Similarly, little difference is seen between male and female common marmosets in their exploration of a new environment, but females show more behavioural activities over all, including touching and manipulating novel objects subsequently placed within the environment (Box, 1988). Subadult female cotton-top tamarins are both the first to explore, and most frequent to explore, a newly accessible area in a captive situation (McGrew & McLuckie, 1986).

Reports that female callitrichids are more responsive to novel food tasks, objects and situations contradict those of Reader & Laland (2001), who found that male primates are more innovative than females. A positive response to novelty is a factor in innovative behaviour, as avoiding newness precludes innovation. Interestingly, when these data are subdivided, males from dimorphic species are indeed more innovative, but in monomorphic species the trend is reversed and females are the more innovative sex (Box, 2003). Thus from exploring innovation, we might expect that female marmosets are the more responsive sex.

1.3.4 Age

Age also affects response behaviours in several species of callitrichids. The order in which individuals in a group of saddleback tamarins approach stimuli, and general response to a novel object, is predictable from the ages of individuals, ranging from the least responsive adult breeding pair to the most responsive juvenile male offspring

(Menzel & Menzel, 1979). Conversely, the older offspring in a cotton top tamarin group respond more quickly and for longer to a novel object than the dominant adults and younger offspring (Millar *et al.*, 1988).

Marmosets also show variation in response due to age. Sub-adult common marmosets explore a novel environment more than adults and juveniles (Rogers, 1999), and older offspring (subadult-adult) tend to touch novel objects more quickly and for a longer time (Millar *et al.*, 1988). Adults and the youngest animals in a group, however, are most responsive during normal feeding in both common marmosets and black tufted-eared marmosets (Box & Smith, 1995). These studies suggest two U-shaped curves of response going in opposite directions, where older offspring at a subadult to adult age are more responsive to novelty but less responsive to familiar food than older and younger animals. This variation in response to novelty may have life-history consequences. In a simulated dispersal experiment, sub-adult females and then sub-adult males were the first to explore newly accessed areas, and spent the most time in them (McGrew & McLuckie, 1986). If older offspring are at an age where they may normally emigrate from a home group, an increased response to new stimuli, environments and individuals would be an advantage.

1.4 Predictions and Hypotheses

In summary, variation in response to novel objects could have several independent or collectively contributing causes. The sex and age of an animal can affect how it responds to a novel stimulus, as might genetically based individual variation, possibly related to a behavioural strategy. Behavioural strategies could also be based on physical aspects of the animal such as size.

In common marmosets, there is some evidence to suggest that females are more responsive to novelty than males, especially when food is involved. Also, when species are monomorphic like the common marmoset, females tend to be more innovative, which is related to responsiveness. It can be predicted from these results that in simple stimulus presentation tests, females will be more responsive than males. Of different age groups of animals, older subadults are the most responsive to novelty, but adults and the youngest animals in a group are most responsive during normal (non-novel) feeding.

The present study addresses the following hypotheses:

1. Common marmosets display individual variation in responsive to novel stimuli that has a measurable genetic influence

2. Sex affects response to novel stimuli in common marmosets, with females being more responsive than males

3. Age affects response to novel stimuli in common marmosets, with older subadults being the most responsive

4. Weight affects response to novel stimuli in common marmosets

1.5 Thesis Outline

The present study aims to investigate the variability of response behaviours in the common marmoset. If variation is found in the trait, it aims to establish whether or not the variation has a genetic basis (i.e. is heritable), or is due to the sex or age of individuals. Chapter Two describes the methodology for the stimulus presentation tests, the nine novel stimuli presented, the study animals used and the behavioural measures taken. Chapter Three describes variation in response using five behavioural measures; latency to approach, latency to contact, duration of proximity, duration of contact and visual attendance. Variation in the five variables is described for each of the nine stimuli presented. This variation in response is then used to classify the stimuli into four groups. Chapter Four examines the effect of sex, age and weight on response. Chapter Five describes the use of Principal Component Analysis to derive general response continua from the range of measures recorded. Chapter Six investigates the heritability of the general response continua and the five behavioural measures. The final chapter, Chapter Seven, provides a discussion of the results obtained, relating the findings to marmoset behaviour and ecology.

Chapter Two

Methods

2.1 Study animals

2.1.1 Study species: common marmosets

The common marmoset (*Callithrix jacchus*) is a diurnal, arboreal platyrrhine primate, native to the Atlantic rainforests of Brazil. Distinguishing characteristics include large white ear tufts and a tail with alternating dark wide and pale narrow bands (Hershkovitz, 1977, p517 for "sub-species" *C. jacchus jacchus*). In the wild, the common marmoset's diet consists mostly of fruits and plant exudates (marmosets eat the highest proportion of plant gum in their diet of any monkey) and also small animal prey (Ferrari & Lopes Ferrari, 1989).

Common marmosets live in complex and dynamic polygynandrous social groups, most groups containing more than one adult of either sex (Ferrari & Lopes Ferrari, 1989, Rothe & Darms, 1993). Not all animals in the group are necessarily related to the breeding male (Nievergelt *et al.*, 2002). Some care giving is given by non-parents, although this is not necessarily required for infant survival (Rothe & Darms, 1993), and lactation is more energy intensive than such behaviours as carrying (Nievergelt & Martin, 2001)

Yamamoto (1993) gives ages as: infant from two to three months, juvenile from five to ten months; and sub-adult from 10-15 months, moving from there into adulthood. Slightly different divisions are given in Hearn (1987, Table 37.1): weaning at 40-80 days; puberty at 8 months; the reaching of adult weight at 12-15 months, but not sexual or social maturity until18-24 months. In captivity, animals can breed successfully up to the age of 14 and beyond, and can live for around 20 years (Box & Hubrecht, 1987). Birth weights have been recorded as 25-35g (although a range from 26.7-39.6g was recorded in this study), with an adult weight range from 386-493g for males, and 382-600g for females (Poole & Evans, 1982).

Marmosets are excellent candidates for behavioural research in captivity (Stevenson, 1977). Due to their small size and the relatively stress-free way in which they interact with humans and react to novelty responsiveness tests (Blackwood, 2000) they are ideal for a large scale study of behavioural responsiveness in a non-human primate. Also, their behavioural repertoire has been noted in detail (Stevenson & Poole, 1976, Voland, 1978). There are, however, several aspects of their biology that it is important to be aware of.

Marmosets usually give birth to dizygotic twins (Sussman, 2002), and occasionally triplets (more often in captivity). Offspring who share a womb exhibit somatic chimerism (Haig, 1999), with cells from both (or all) products of conception colonising the bone marrow of both (or all) the developing individuals. This is important in genetic studies of common marmosets as it means that blood samples cannot be reliably used, and tissue or hair samples are required for genotyping. This does not affect heritability studies, as all siblings can be assumed to be dizygotic.

Along with the majority of South American primates, marmosets are polymorphic for colour vision, all males being dichromatic, and females dichromatic or trichromatic, depending on the homozygosity of a single autosomal locus (Jacobs, 1998). This will lead to variation in individual's ability to perceive colours (Caine & Mundy, 2000), and possibly discern colour-based differences between stimuli.

2.1.2 The animals

The study group of 68 common marmosets consisted of 26 males and 42 females. This sample was taken from a breeding colony population at DSTL Porton Down, Salisbury, where tests were carried out between January and December 2002. All the animals had been removed from their family groups at 7-10 months and were living in same sex peer group gang cages. Age at the commencement of testing ranged from 14 months to 22 months (see Table 2.1). Group size ranged from seven to 13, with one group (seven) being split into a three and a four during testing, because of intra-group aggression. A table of all individuals' sex, age at testing, birth weight and test weight, and litter size can be found in Appendix One. All data were recovered from colony records. These measures were important in investigating whether responsiveness was affected by the sex, age or size of the animals. It is possible that a greater litter size may have a

Table 2.1 Descriptive statistics for the 68 individuals in the study sample.

	Birth weight (g)	Test weight (g)	Litter size	Age at testing (days)	Age at testing (months and days)
Mean	32.63	334.13	2.48	442.53	14m14d
Minimum	26.7	271	1	322	10m1d
Maximum	39.6	473	4	673	22m29d
Range	12.9	202	4	351	12m28d
Standard deviation	3.39	39.94	0.61	92.88	

constraining effect on the weight of the animals. There was a small but significant negative correlation between litter size and birth weight ($r = -0.383$, $p=0.001$), but this disappears by the age of testing ($r = -0.044$, $p=0.723$). Birth weight and the weight at testing are significantly correlated ($r = 0.345$, $p=0.004$), but not very highly, indicating that the animals born bigger tend to remain bigger than their conspecifics. There was no difference between sexes in weight at birth ($T=-0.39$, $df = 49$, $p=0.70$), but by the time of testing, females were significantly heavier than males ($T=2.09$, $df= 65$, $p= 0.040$). There was no difference in litter size for males and females ($T=-0.40$, $df= 49$, $p=0.69$).

2.1.3 Colony housing and husbandry

The breeding colony rooms had a light-dark cycle of 14 hours light (05:45-19:45), including a half hour dawn, and 10 hours of darkness, including a half hour dusk. Temperature was maintained at around 23.5°C and humidity at 55%. The animals were fed once daily, at approximately 2pm. Exact feeding time varied slightly from day to day during the study depending on both husbandry and the duration of testing on that day. Feeding was *ad libitum*, with primate diet (Special Diet Services, Witham, Essex, UK), apples and oranges being given daily, supplemented by bananas, grapes, pears and raisins once a week. Once weekly the animals were given malt loaf, and on a different day "milkshake mix", comprising of baby rice, SMA milk powder, Complan, glucose powder and vitamin D3. In addition to this, boiled eggs were given twice weekly with an added SA3-7 vitamin supplement. Food was given in large or small sized non-reflective stainless steel trays, dependent on the number of animals per cage. Fresh water was supplied in bottles attached to the cage sides, changed daily at approximately 10am. The bottles themselves were changed and washed weekly. Trays at the bottom of

the cages were laid out with sawdust and small items of food for foraging, such as assorted nuts (including monkey nuts), freeze dried bananas, rice crispies and raisins. (Information supplied by animal husbandry staff).

Animals were housed, depending on group size, in either single gang (i.e. peer-group) or interconnected gang cages. Cages were of two sizes. Small cages measured width 1220mm, depth 520mm, and height 1620mm, with a 610 by 520mm shelf 760mm below the cage top (holding a maximum of four animals). Large cages, used for both families and some single sex groups, measured width 1550mm, depth 850mm and height 1620mm, with a 770 by 760mm shelf 760mm from the cage top (holding a maximum of six adults). Multiple cages for larger groups were connected by extendable semi-transparent plastic and metal "elephant trunk" tubing (15cm diameter). Metal

Figure 2.1: An example (small size) cage, showing attached pre-test area cage on the left, and elephant trunk tubing on the right. The buckets, perches and terracotta coloured plastic tubes used as enrichment can be seen in the cage. A water bottle is attached to the top left quarter of the cage. The shelf used for stimulus presentation has a plastic tube on it.

slides could be fitted over the entrances to the cages, restricting animals' movement into the connective tubing and hence other cages. Cages were equipped with wooden perches for the animals to climb on, gnaw and scent mark. The perches were replaced on a weekly basis. In addition, enrichment was provided by rope and wood ladders and swings, Perspex platforms, terracotta coloured plastic tubes (approximately 30cm long and 10 cm in diameter) and white plastic buckets for the animals to climb on and in. The buckets were also used for sleeping in. Figure 2.1 shows an example cage, with the pre-test area cage attached (see 2.2.2).

A minimum number of cages were required for the test procedure (the exact number depending on group size), so some groups had cages added or rearranged before testing took place. The test set up is described in Figure 2.2. If cage set ups were altered to achieve this design, a four day minimum period of acclimatisation for the animals was allowed before any testing took place. The arrangement of the tubing and the connections between cages were retained as constant throughout the testing. Outside test period, the post- and pre-observation cages shown in Figure 2.2 were connected.

2.2 Presenting stimuli

2.2.1 Training and habituation

Testing was based on the presentation of novel stimuli to individual animals when they were temporarily isolated from their group, but still within a home cage area. Animals were moved one at a time from a pre-observation cage into an observation cage (Figure 2.1), where the response to a stimulus was recorded over a 240 second presentation. Animals were then moved into a post-observation cage (Figure 2.2). To this end, the animals needed to be habituated to the presence of the observer and the mechanics of the testing. Animals were exposed to the observer working in the room with other marmoset groups and aiding with husbandry (e.g. feeding) before testing. Depending on the size of groups and their cage sizes and position in the colony, slight variations in the standardised training and habituation occurred by necessity. The description of the basic training procedures for each group of animals tested is presented below.

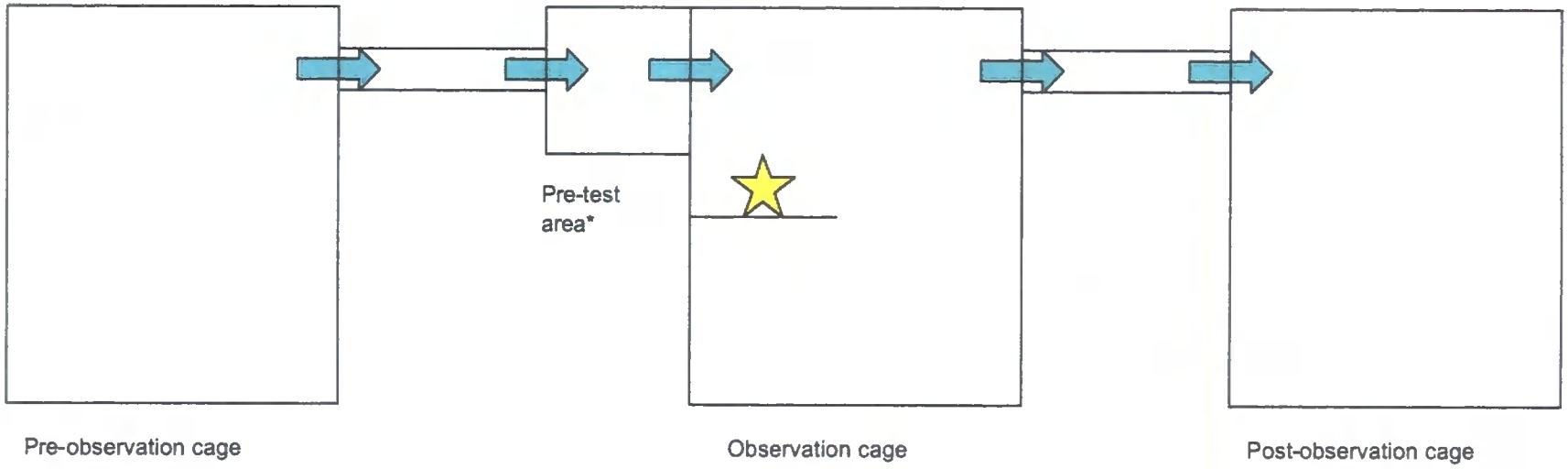
The initial task for the observer each day was to gather all of the animals in the group into the pre-observation cage (see Figure 2.2). This was achieved simply by the observer moving over to that cage on entry into the colony room. Almost all animals

would approach the observer, who would then close the metal slides at the cage exits, thus restricting the animals' movements. Any animals remaining in other cages could then be individually moved into the pre-observation cage by tapping on the cage and tubes to attract the animal, or gentle encouragement (e.g. opening the cage and moving an arm towards the animal until it moved into the connective tubing), using the metal slides as doors to stop other individuals leaving. This process had to be repeated every morning prior to testing. Once in the pre-observation cage, animals were let one at a time, using the metal slides to control access, into the observation cage, and then into the third, post-observation cage, until all had passed through. Animals were not selected by the observer, rather by opportunistically allowing access to individual animals as they approached. This was repeated over several days, with one or two run-throughs a day. On the fourth run-through, each animal was left in the observation cage for a minimum of one minute before being let out. After all animals passed one at a time through the set up five consecutive times without signs of undue stress, a test presentation with a novel object was given. The object, a green plastic wine goblet, had been assessed as non-threatening in previous tests. Once all animals were exposed to a four-minute presentation period with the goblet without exhibiting signs of fear or stress, the novel presentation testing proper was started the next working day. Presentations in the test cage were not visible to other animals and stimuli were kept hidden from all animals in the colony when not used. Animals were thus completely naïve to the goblet and all other novel objects before their own test, and visual social interactions during the test procedure were avoided.

2.2.2 Novel stimulus presentation testing

In the test series each animal was presented with an identical set of nine stimuli (Table 2.2, Figure 2.3), one stimulus per test run. Individual objects were not repeatedly presented to subsequent animals to avoid any olfactory cues from previous subjects (except the mirror, which was disinfected and rinsed between each use). The observation cage had a "pre-test area" metal box attached to the entrance where animals could be identified before entering the main test area, and presentations of the various stimulus types were randomised for each animal to avoid order effects across the test days. All objects were placed in the same position in the centre of the cage shelf except the mirror, which was leant against the inside wall of the cage on the shelf. After animal identification and the placing of the stimulus in the cage, the observer opened the metal

Figure 2.2: Home-cage set up for test presentations (not to scale).



Sealable entrance/exit, showing direction of movement through test

Stimulus on shelf

Connective tubing between cages

* For identification of animal prior to presentation. Other areas are standard home cages.

Observation cage
width 1220cm, depth 520mm, height 1620mm; with a 610mm by 520mm shelf 760mm below the cage top

Pre- and post-observation cages
width 1550mm, depth 850mm, height 1620mm; with a 770mm by 76omm shelf 760mm from the cage top

Outside of testing times, pre and post observation cages are connected

slide separating the pre-test area and presentation cage. Recording started once the animal had moved its head through the entrance into the observation cage. Presentations lasted 240 seconds. This time was chosen based on previous work that indicated in this kind of test set up, if an animal is going to approach and manipulate an object, it will do so within four minutes of entering the cage (Blackwood, 2000). All stimuli were immediately visible to a subject entering the cage. The observer was present during all presentations, allowing the animal into the observation area using the above procedure, and then standing in a fixed position approximately three metres from the cage.

After the presentation test the animal was let into the post-observation cage, via the connective tubing, by opening the metal slide over the exit (after the presentations involving food rewards, RIT and RIW, if an individual had not retrieved a raisin it was given one by the observer). The next animal would then be let into the pre-test area and the procedure was repeated until all animals in the group had been presented with a stimulus.

2.2.3 Stimulus selection

The objects presented (Table 2.2, Figure 2.3) were selected because of a combination of suitability and availability. One stimulus of each type was required for every member of a group tested. This was so no object was presented twice during a session, avoiding potential olfactory cues from previous animals. Scent marking and olfactory communication are important in callitrichids (Epple *et al.*, 1993). All objects were disinfected before each use (including the first time every object was presented), again to avoid olfactory cues.

A number of preliminary trials, including previous work with 13 individuals (Blackwood, 2000) and tests on 22 animals within the current research project, were carried out to assess the suitability of different stimuli for object presentation tests. None of these animals were in the final 68 used for the main presentation tests. The nine stimuli used were chosen because of the variation in responses from these preliminary tests.

Within the group of stimuli chosen, two have possible "social" aspects to them. The mirror (MIR), obviously having the potential to reflect the subject toward itself and the

mouse and cheese toy (MOU) having a protruding head and facial features approximately the same size as a marmoset's. Two other stimuli included food rewards. RIT, raisin in tube, had two irretrievable raisins. The two raisins in RIW, raisin in water, were obtainable but submerged. In different colonies, both scientists and husbandry staff have used and tested food stuffs as various as cream cheese, yoghurt, liquorice, tuna fish, pepper, raisins and nuts as rewards and aversive foods during testing and also in enrichment. Raisins, an easily available preferred food used at Porton, were decided as most appropriate in these tests.

The five remaining stimuli (TUB, a grey tube; EGG, a cardboard egg tray; POT, a blue pot; SYR, a syringe; and BAL, a hollow pink ball) were chosen because of the interest shown by animals in the preliminary presentation trials. These objects were all manipulable, with holes for the animals to reach into or protrusions to hold or bite. As the animals could lift or move the objects, this manipulability may be related to the controllability of a stimulus, which can be rewarding for an animal in enrichment situations (Sambrook & Buchanan-Smith, 1997). Stimuli that are attractive to animals because of such qualities can be more useful in assessing individual differences in novelty responsiveness. This is because if an object is attractive the more responsive animals should approach it quickly, and the less responsive animals (or animals responding negatively) approach slowly, if at all. If an object is not stimulating, it is possible that none of the animals will approach it. If objects are previously known to stimulate responsiveness then it is more likely that it is the reaction of the animal rather than the attractiveness of the stimulus that is being assessed. There may well of course still be variation between potentially attractive objects in how animals respond to them. Conversely to the other stimuli, MOU, the mouse head and cheese toy, had elicited negative (rather than no) reactions in preliminary tests. For this stimulus then, it is not interest versus disinterest being investigated, but negative versus positive interest. As Figure 2.1 and Figure 2.3 illustrate, the novel stimuli selected were not similar to the enrichment objects or elephant trunk present in the cages on a daily basis.

The stimuli presented all represented novelty in some way and thus could be used to assess variation in response. Dissimilarities between the stimuli meant that different aspects of responsiveness could be examined. For some stimuli, such as MIR, RIT and RIW, there was an opportunity to assess different behavioural traits related to response.

Table 2.2 Novel stimuli used in presentation tasks

Stimulus Code	Description
MIR	Mirror (150 by 150mm, 50mm deep back)
TUB	Corrugated grey plastic tube (approx. 130mm length, 20mm diameter)
MOU	Mouse head in cheese rubber dog toy (60 by 80mm, with a 50mm across protruding head)
EGG	Cardboard egg tray (140 by 140 by 40mm)
RIT	Raisin in sealed transparent 50ml tube
RIW	Raisin in water filled metal food dish (125 by 150 by 25mm)
POT	Sealed blue plastic pot (75mm by 60mm)
SYR	Transparent 60ml plastic syringe
BAL	Hollow pink plastic ball with holes (approx. 60mm diameter)

Figure 2.3: Stimuli for presentation. Top row L-R: Egg tray (EGG), Mirror (MIR); middle row L-R: Tube for raisin (RIT), blue pot (POT), pink ball (BAL), mouse and cheese toy (MOU); bottom row L-R: syringe (SYR), corrugated grey tube (TUB). The standard metal food dish of the raisin in water stimulus (RIW) is not included.

2.3 Data recording and analysis

2.3.1 Recording

Data were recorded by direct observation using all occurrences sampling of response behaviours. The range of behaviours noted, including approach, sniffing and touching, are given in Table 2.3. An example of the check sheets used can be found in Appendix Two. The data recorded were used to create four measures of responsiveness:

(1) Latency to approach within one body length (not including tail) of the object

(2) Latency to contact (touching with nose, mouth or hands)

(3) Duration of proximity (within one body length)

(4) Duration of contact (touching)

During presentations, all the behaviours recorded were events. Duration times were calculated using event times, e.g. the time between approach stimulus and move away from stimulus was used to calculate the duration of proximity to the stimulus. All proximity duration times for each presentation session were added together to give one overall score, e.g four five-second periods of close proximity to the stimulus would give a 20-second duration of proximity score for that presentation. The same was carried out for duration of contact times. In addition to these measures, every 10 seconds it was noted whether the animal's head was oriented towards the object. From this, a fifth measure (5) "the proportion of time spent visually attending to the stimulus" was calculated. These five measures are not independent of each other, but may cover different aspects of an animal's response to a stimulus. Autogrooming, scratching and "wet dog shaking" were recorded as possible measures of stress (Stevenson & Poole, 1976; Barros *et al.*, 2000). They were not seen during the testing often enough to be statistically analysed, and so have not been included in any subsequent analysis.

2.3.2 Statistical analysis

Statistical analyses were carried out using MINITAB (regression analyses, ordinal logistic regression, MANOVA, t-tests; MINITAB release 13.1, MINITAB Inc.) and SPSS (ANOVA, principal component analysis and post-hoc Sheffe tests; SPSS for Windows, Rel. 11.0.1., 2001), except Bonferroni multiple comparison calculations, which were carried out by hand from Sokal & Rohlf (1994), and Power analysis, which was carried out using G*POWER (Erdfelder *et al.*, 1996; Buchner *et al.*, 1997).

Table 2.3: Behaviours recorded during stimulus presentation tests. All behaviours were recorded as events, and durations calculated subsequently (see 2.3.1)

Behaviour	**Description**
Approach stimulus	Animal moves to within one body length of the stimulus with its head oriented toward it.
Sniff/Touch stimulus with nose	Animal sniffs the stimulus or touches it with its nose. This can only occur when the animal is within one body length of the stimulus.
Manipulate stimulus	Animal manipulates the stimulus with its hands, feet or mouth (biting rather than sniffing) whilst oriented towards it. This can only occur when the animal is within one body length of the stimulus.
Stop manipulating stimulus	Animal stops touching the stimulus with its hands, feet or mouth, but remains within one body length of it. This behaviour only occurs subsequently to manipulate stimulus.
Move away from stimulus	Animal moves from within one body length of the stimulus to further than one body length away. This behaviour can occur subsequently to approach stimulus or manipulate stimulus.
Jump towards stimulus	Animal moves to within one body length of the stimulus whilst oriented towards it, with a fast, bounding gait where all four limbs leave the ground.
Jump away from stimulus	Animal moves away from the stimulus with a fast, bounding gait where all four limbs leave the ground.
Autogroom	Animal moves its hands through its body hair, often visually inspecting the area at the same time.
Scratch	Animal moves one hand through its own body hair in fast, repetitive movements.
"Wet dog shake"	Animal shakes quickly, like a "wet-dog" (Barros *et al.*, 2000). This behaviour could occur if the animal is wet (stimulus six).

Pedigree construction was carried out using PedSys (PedSys Version 2.0; Dyke, 1989). The computer package Sequential Oligogenetic Linkage Analysis Routines (SOLAR version 1.6.6; Almasy and Blangero, 1998) was used to estimate heritabilities. As statistical treatment differs in each chapter, detailed descriptions of analyses are given individually for each one.

2.4 Welfare during testing

The welfare of the individuals during testing was important, not only for the animals themselves, but for the accuracy of the recording. If an animal was stressed before entering the observation cage, behaviours recorded would not be an accurate measure of response to the stimulus. All animals tested were habituated to both the presence of an observer and the mechanics of the testing (see 2.2.1) before any presentations were given. A subjective assessment of whether or not to test the animals was taken every morning on entry into the colony. If the animals were physically fighting, or bullying an individual to the extent that the group could not all be in the same cage (required due to the mechanics of testing), testing was not carried out. Also, if individuals alarm called excessively and chronically, or darted continuously about the cage in reaction to the observer's or another animal's presence, testing was not carried out. Noise levels in a facility can affect an animal's behaviour (Milligan, *et al.*, 1993), and if noise from other cages (such as alarm calls and scuffles) or husbandry elsewhere in the colony was significantly disrupting the behaviour of the test animals testing was also stopped. Based on personal observations, in such situations animals will not respond to novel stimuli, being more interested in what is happening outside the test cage. Within the observation cage, buckets and tubes allowed the animals to avoid visual contact with the stimulus (and the observer) if desired. Also, the bottom corner of the cage under the shelf was out of sight of the stimulus. If animals became agitated during a presentation, observation was continued for the 240-seconds, and then testing was stopped until animals had settled, or until the next day. If the testing had been stopped mid presentation, then the stimulus would no longer have been novel when re-presented. The human observer was present at all times, and could allow the animal out of the cage if it became unduly stressed (indicated by fast darting about the cage and/or constant alarm calling).

Chapter Three

Measures of Responsiveness

3.1 Introduction

Five variables representing responsiveness were measured during presentations: Visual attendance; Latency to first approach, Latency to first contact; Duration of proximity and Duration of contact. These variables are not independent of each other for several reasons. Firstly, some measures are dependent upon the occurrence of others. Latency to first contact is dependent on latency to first approach, as an animal cannot touch a stimulus without first coming within one body length of it. Similarly, duration of contact is "nested" within duration of proximity, again because to touch a stimulus the animal must be in proximity to it. Secondly, visual attendance will be expected to occur when an animal is in proximity to or touching a stimulus, as it is more likely to be paying attention to it. Visual attendance then is an overview measure of response that includes but is not dependent on the periods of physical attention measured by the other variables. There are thus hierarchical clusters of interdependent measures. All duration and latency times are intertwined, with only latency to approach and visual attendance not being constrained in some way by the other measures.

This chapter presents the results for the five measures taken, grouped by mean scores for each animal, and by stimulus. The chapter assesses the suitability of both measurements and stimuli for use in subsequent investigation of individual variation in responsiveness. It is important that there is enough variability in each measure to be able to detect individual differences. This is necessary for the stimuli presented as well. Additionally, it is possible that different measures will be the most appropriate to use for certain stimuli, depending on the characteristics of the stimuli themselves. Sections 3.3 to 3.7 assess each of the five measures used individually and sections 3.8 and 3.9 discuss using these results to characterise stimulus response and organise the data for further analysis.

3.2 Analysis

Throughout the chapter, box and whisker plots are used to describe the data. Interquartile ranges (IQRs), indicated by boxes in the diagram (e.g. Figure 3.1), give a measure of spread of the data that is unaffected by anomalous outliers (see Figure 3.1 legend for detail).

The statistical tests used in analysis were different depending on whether the data were normally distributed. Normality was tested for both mean response scores and response scores for each stimulus using Kormogorov-Smirnov tests (SPSS for Windows, Rel. 11.0.1; 2001). Log transformations were used to normalise data distributions when possible. If the nine stimulus scores for a response were normally distributed, analysis of variance (ANOVA) and Sheffé post-hoc tests were used to investigate significant differences in the response between the stimuli. Where the data were not normally distributed, Friedman's analysis of variance by ranks and related post-hoc comparisons (Siegel & Castellan, 1988) were used to assess stimulus differences.

Latency times were analysed in two different ways due to the treatment of censored (missing) data, when an individual did not approach the stimulus within the four-minute presentation period. Firstly, censored data were removed, and the remaining time scores for each individual and stimulus assessed. The number of censored scores varied between stimuli, due to general differences in reaction, as seen for other measures. Secondly, latency scores were calculated by converting times and censored data into categorical data to investigate the reactions of all animals to all stimuli. Scores run from one to four, with animals quickest to respond scoring one, and four being the censored values, so assigned to those animals that did not approach the stimulus (see table 3.5).

3.3 Visual attendance

Visual attendance is the most general measure of responsiveness used, recorded by noting the orientation of an individuals head (toward or away from the stimulus) every 10 seconds during a four minute presentation. As it does not rely on proximity, unlike the other measures, it is not necessarily a "positive" reaction to a novel stimulus, merely a reaction. This could for instance include fixation on a threatening object. The data

have a mean of 0.63 and a standard deviation of 0.18 (range 0.21 to 0.89). This mean falls to the right hand side of the possible distribution, showing individuals on average spent more than half of the presentation time visually attending to objects. Most animals scored between 0.45 and 0.9, with three animals falling below this. These results are slightly left skewed, but a one sample Kormogorov-Smirnov (K-S) test shows them to be normally distributed (K-S Z= 0.597, p= 0.868).

3.3.1 Visual attendance scores by stimulus

Using individual mean scores to investigate the data gives a general impression of how the animals respond to the stimuli presented. It is also possible to look at each stimulus individually, to understand how reactions to them differ. Figure 3.1 shows visual attendance for each stimulus. Differences between the stimuli are apparent. All data ranges overlap to a lesser or greater extent, and the relative centrality of the medians in the IQR box indicate little skewness except for RIT, the irretrievable raisin in a sealed

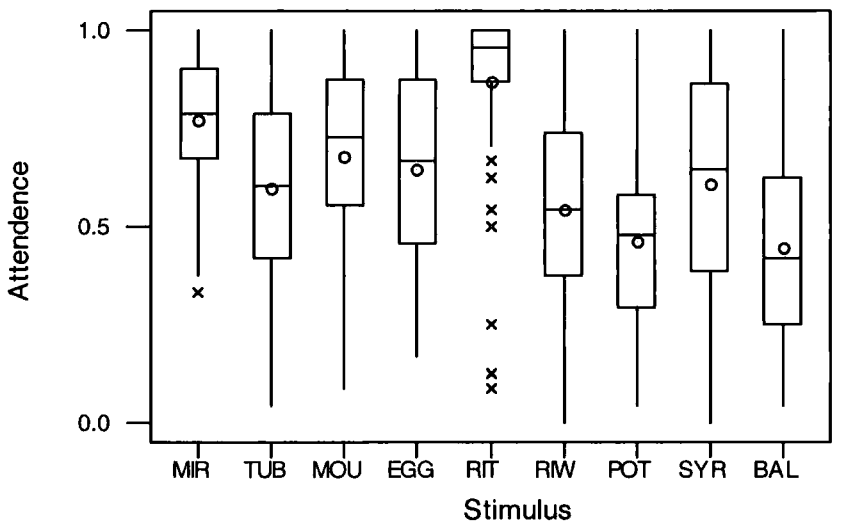


Figure 3.1: Visual attendance to nine novel stimuli for 68 individuals. The x-axis shows stimulus code (see Table 2.2), the y-axis the proportion of time spent in visual attendance. Circled dots indicate the mean score for each stimulus, crosses indicate outliers. Boxes indicate the range between the first (Q1) and third (Q3) quartile, bisected by the median. Whiskers indicate upper and lower limits (lower limit: Q1-1.5 (Q3-Q1), upper limit: Q3 + 1.5 (Q3-Q1)).

tube. This stimulus shows both the smallest IQR (0.126) and the highest median (0.96) and mean (0.87). This indicates that the majority of the animals spent almost the whole presentation period visually attending to the stimulus. There are however, several outliers; three individuals spent less than 50 percent of the presentation period attending. The common high level of attention may reflect the presence of food in the stimulus. RIW, a raisin food reward under a layer of water, receives considerably less visual attendance, with median and mean of both about 0.54. The larger IQR (0.36) of RIW also shows a greater range of scores than for the other food stimulus. Of the stimuli without food rewards, the mirror, MIR, has the highest median (0.79) and mean (0.77) scores, the second highest measures of central tendency overall after RIT, and the second lowest IQR overall (0.28), again after RIT.

Of the other six stimuli, only two, POT (the blue pot) and BAL (the pink ball) had both means and medians below 0.5. Visual attendance for POT had a mean of 0.45 and a median of 0.47, for BAL a mean of 0.44 and median of 0.41. RIW and SYR, the plastic syringe, both had minimum scores of 0, indicating a lack of overt response on the part

Table 3.1: ANOVA post hoc comparisons (Sheffé) for differences in visual attendance between stimuli. The top number refers to the mean difference (I-J, where I is the column stimulus and J the row stimulus. The bottom number refers to the significance (p-value). Results significant at p=0.05 are in **bold**. Comparisons for RIT are outlined as an example

Stimulus	MIR	TUB	MOU	EGG	RIT	RIW	POT	SYR
TUB	**0.1728** **0.014**							
MOU	0.0919 0.704	-0.0809 0.833						
EGG	0.1232 0.277	-0.496 0.991	0.0312 1.000					
RIT	-0.0999 0.594	**-0.2727** **0.001**	**-0.1918** **0.003**	**-0.2230** **0.001**				
RIW	**0.2304** **0.001**	0.0576 0.976	0.1385 0.134	0.1072 0.487	**0.3303** **0.001**			
POT	**0.3137** **0.001**	0.1409 0.117	**0.2218** **0.001**	**0.1906** **0.003**	**0.4136** **0.001**	0.0833 0.807		
SYR	**0.1648** **0.0.25**	-0.080 1.000	0.0729 0.902	0.0417 0.997	**0.2647** **0.001**	0.0656 0.946	-0.1489 0.074	
BAL	**0.3033** **0.001**	**0.1575** **0.043**	**0.2384** **0.001**	**0.2071** **0.001**	**0.4301** **0.001**	0.0999 0.593	0.0165 1.000	**0.1654** **0.024**

of at least one animal. Conversely, all stimuli had maximum scores of 1, indicating that at least one individual spent all the available presentation time visually attending to the objects. For all stimuli then, the range of scores covers almost all possible times spent in visual attendance, MIR and RIT having the highest measures of central tendency and the smallest IQRs. The largest IQRs are seen for SYR, and EGG, the egg tray. They do not, however, have the lowest mean or median score, so there is no obvious relationship between these scores and IQR size across the stimuli.

An analysis of variance (ANOVA) demonstrates that the variation seen in Figure 3.1 between stimuli is significant ($F_{(8,603)}= 26.117$, $p< 0.001$). Post hoc tests show that the main difference is between RIT and the other stimuli, and that RIT is significantly different to all others except MIR (see Table 3.1). Other differences are highlighted in the table.

3.4 Duration of proximity

Figure 3.2 shows the spread of mean proximity duration scores for the nine stimuli over the 68 animals measured. The data have a mean of 92.0 and a standard deviation of 34.6 (range 10.8-165.1). As the maximum theoretical score possible is 240 seconds, scores are not spread evenly across the presentation time. This could be because of an initial

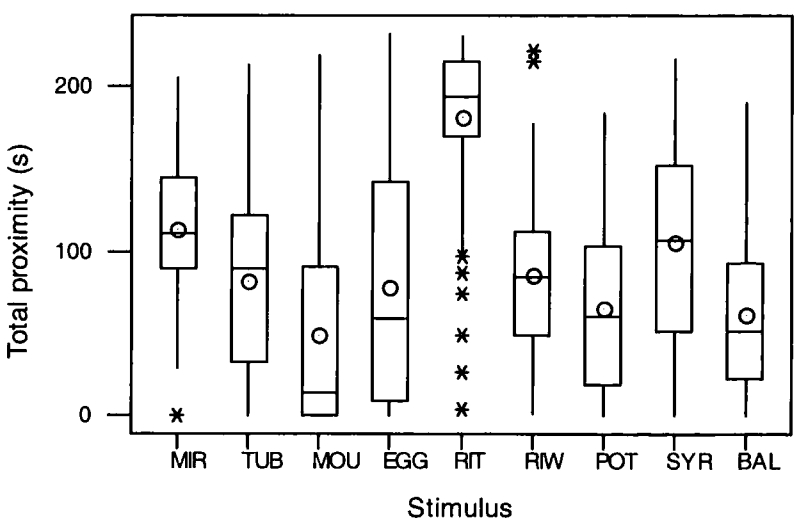


Figure 3.2: Total time spent in proximity to nine novel stimuli for 68 individuals. The x-axis shows stimulus code (see Table 2.2), the y-axis the total time spent in proximity to a stimulus. Symbols as for figure 3.1.

delay in approaching stimuli, assessed below as latency to approach scores. A one-sample K-S test shows that the data are normally distributed (K-S Z=0.633, p=0.817).

3.4.1 Duration of proximity for each stimulus

As with visual attendance, it is possible to look at proximity duration scores stimulus by stimulus. Figure 3.2 demonstrates variation in response using box and whisker plots. Unlike for visual attendance, none of the stimuli have a range that encompasses the entire presentation time. Even accepting that an animal may need several seconds to reach the stimulus to be within one body length of it, no animal is in proximity to any stimulus for more than 232 seconds. There are always at least 8 seconds during a presentation when an animal is not in proximity to the stimulus. IQRs, means and medians do demonstrate that most animals had much lower proximity duration times than this. All stimuli except POT and BAL, however, have reasonably wide variation in response to them, ranging over 200 seconds. The smallest IQR (46.25) and highest median (194) and mean (181.3) are those of RIT, the irretrievable food reward in a sealed tube, demonstrating the smallest variation in response. Second to this was MIR,

Table 3.2: Friedman's ANOVA by ranks post hoc comparisons (Siegel & Castellan, 1988) for differences in proximity duration between stimuli. Significant differences are marked by an asterisk in the bottom left part of the grid. The minimum difference in mean rank required for significance is 1.5. This significant difference is calculated as $\geq Z\alpha_{/k(k-1)} \sqrt{k(k+1)/6N}$, where $\alpha = 0.05$, N= sample size (68) , k= number of conditions (9). Z can be calculated using appendix AI of Siegel & Castellan (1988).

Stimulus (and mean rank)	MIR 6.21	TUB 4.76	MOU 2.96	EGG 4.20	RIT 8.54	RIW 4.89	POT 3.88	SYR 5.72
TUB 4.76								
MOU 2.96	*	*						
EGG 4.20	*		*					
RIT 8.54	*	*	*	*				
RIW 4.89			*		*			
POT 3.88	*				*			
SYR 5.72			*	*	*		*	
BAL 3.85	*				*			*

also having a relatively high median (110.5) and mean (114.1), and a low IQR (56). All other stimuli except SYR had both a mean and median of less than 100 seconds. Conversely to MIR and RIT, SYR showed an IQR of 101.5, second highest after EGG, with an IQR of 135.5. MOU had both the lowest mean (48.6) and median (13), and the third highest IQR, 91. There is therefore no obvious relationship between average scores and IQR across the stimuli.

As with visual attendance scores there is significant variation between the stimuli (Friedman's Analysis of variance by ranks, N=68, χ^2_8= 201.055 p< 0.000), with RIT having significantly higher proximity scores than any of the other stimuli. Table 3.2 shows the full array of comparisons between individuals. MOU has the greatest difference in scores to RIT, and is significantly lower than five of the other stimuli. MIR has the second highest response scores, and is significantly different to five other stimuli, as is SYR.

3.5 Duration of contact

Duration of contact is a measure of how long during the 240-second presentation an animal stayed in physical contact with a stimulus. Contact included manipulation with

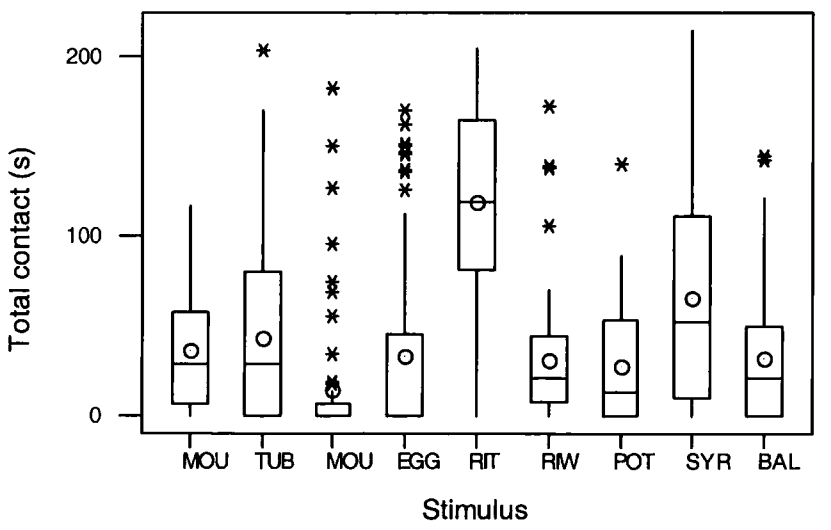


Figure 3.3: Total time spent in contact with nine novel stimuli for 68 individuals. The x-axis shows stimulus code (see Table 2.2), the y-axis the total time spent in contact with a stimulus. Symbols as in Figure 3.1.

hands nose and mouth, or resting on the stimulus with hands or feet, but not brushing past or touching with the tail. As noted above, the measure is dependent upon the animal being close enough to the object to touch it. All scores will thus be lower than the corresponding ones for duration of proximity. The frequency of mean contact duration over all stimuli is shown in Figure 3.6. The data have a mean of 45.2 and a standard deviation of 25.9 (range 0.2 to 103.5), scores being lower, and less spread across time than for proximity duration. The distribution of scores is normal (one sample K-S test; K-S Z=0.916, p=0.314).

3.5.1 Duration of contact by stimulus

Splitting individuals' mean contact duration scores by stimuli shows, as before, some variation in how the animals react to the different objects (Figure 3.3). In general, and as would be expected, scores are lower than for proximity duration, and IQRs are both smaller and much lower in the possible data range. RIT has the highest mean (118.1) and median (118.5), both higher than the third quartile (Q3, top of the box in a box and whisker plot) of the other eight stimuli. In the previous two measures, however, RIT does not have the lowest IQR, but the second highest (83.25), indicating a much wider range of response than for attendance and proximity measures. SYR, the syringe, has

Table 3.3: Friedman's ANOVA by ranks post-hoc comparisons for differences in contact duration between stimuli. See Table 3.2

Stimulus (and mean rank)	MIR 5.06	TUB 4.99	MOU 2.72	EGG 4.07	RIT 8.35	RIW 4.88	POT 4.11	SYR 6.26
TUB 4.99								
MOU 2.72	*	*						
EGG 4.07								
RIT 8.35	*	*	*	*				
RIW 4.88	*		*		*			
POT 4.11	*				*			
SYR 6.26			*	*	*		*	
BAL 4.56			*		*			*

the largest IQR (100.5) and the largest range (214.5). The very low IQR of MOU and the median score of 0, (with a mean of 14.3) shows very little inter-individual variation in a low response, with only three animals of the 68 spending more than 100 seconds in contact with MOU. MIR, EGG, RIW, POT, and BAL all have IQRs of 50 or under, falling at the bottom of the possible data range. Five of the stimuli (TUB, MOU, EGG, POT and BAL) have a first quartile (Q1, bottom of the box in a box and whisker plot) of zero. EGG and MOU also have a Q1 and a median of zero. Overall then, excepting RIT and to a lesser extent SYR, contact durations are low and with little variation in response compared to visual attendance and proximity duration.

As with the previous two measures, there is significant variation between stimuli for duration of contact (Friedman's ANOVA by ranks, N=68, χ^2_8= 191.949, p<0.001). Although MOU and RIT are again the most different, with MOU being the lowest and RIT the highest, the scores of the other stimuli are more evenly spread out between these two. Table 3.3 shows that RIT is significantly different to all other stimuli, and MOU is significantly lower than six others. SYR is significantly different to five others, being lower than RIT, and higher than MOU, EGG, POT and BAL.

3.6 Latency to first approach

Latency to first approach times were analysed in two different ways, Firstly, by removing censored data and assessing the remaining time scores using ANOVA; and secondly, by including censored data and converting time scores into category data. Categories are illustrated in Figure 3.5, one being the fastest response and four being a censored score due to non-approach.

The frequency of mean latency to first approach, when non-approaches within the presentation time are ignored, has a maximum score for any individual of 102.78 seconds and the lowest 12.89 seconds (a range of 89.89 seconds). This demonstrates that if animals are going to touch an object, they will do it within a relatively short space of time, but not immediately on entering the presentation arena. With a mean time over all individual means of 41.6 seconds (standard deviation 21.57) this indicates a slight right skew to the data, but one-sample S-K tests shows them to be normally distributed (K-S Z= 1.209, p= 0.107). A log transformation of the data does move it further away from non-normality (K-S Z= 0.107, p=0.962).

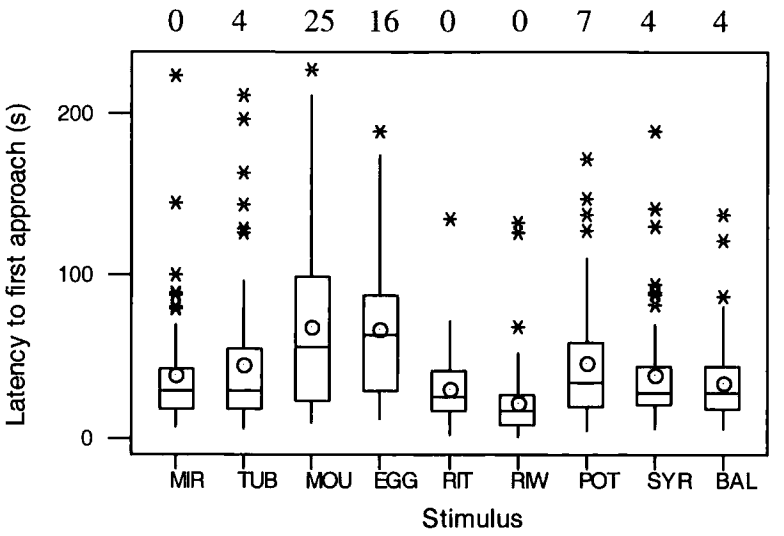


Figure 3.4: box plot of latency to first approach to nine novel stimuli for 68 individuals. The x-axis shows stimulus code (see Table 2.2), the y-axis the latency to first approach time in seconds. Symbols are as for figure 3.1. Numbers across the top of the plot describe the numbered of censored scores for each stimulus.

3.6.1 Latency to first approach scores by stimulus

Latency to first approach times can be assessed stimulus by stimulus, as with duration scores. It should be noted that due to the nature of latency times, a stronger (in this case quicker) level of response gives a lower score. If duration and latency scores are linked it would thus be expected that stimuli with a generally high duration score would have a low latency score. Figure 3.4 indicates that this is indeed the case when compared to Figures 3.2 and 3.3. RIT and RIW, the two food-related stimuli, have both the lowest mean (29.99 and 21.19 respectively) and median (25 and 17) latency scores. Interquartile ranges are quite low, with only MOU, the mouse and cheese toy, and EGG, the cardboard egg tray, scoring above 50 (77 and 58.75 respectively). All stimuli, however, do show outlying individuals taking longer to approach the stimulus. In addition, all except the food related stimuli have censored data, where individuals failed to approach within the four-minute presentation period. The lowest mean, median and interquartile range are all shared by RIW. This being a water-filled food dish, the initial appearance of the stimulus itself will be very familiar to the animals; so this result may be expected. If the animal approaches the stimulus and then notices it is unusual because of the water, the corresponding latency to contacts may be larger. The visual

impression from Figure 3.4 of a relationship between censored value and IQRs, which measure the data spread, is supported by a Spearman's correlation ($r= 0.973$, $n=68$, $p<0.001$). This indicates a simple relationship between animals taking longer to approach a stimulus and animals failing to approach, rather than a quantitative difference between approachers and non-approachers. Log transformation of the data for each stimulus improves normality in all cases, so the transformed data were analysed with ANOVA for significant differences. Differences between stimuli in latency to first approach are significant (ANOVA $F_{(8,543)}= 15.06$, $p> 0.001$), with latency to first approach to RIW being lower than for all other stimuli, including RIT (Table 3.4). RIT is different to only three of the other stimuli, MOU, EGG and RIW. Correspondingly, latency times in response to MOU and EGG are significantly greater than three and four of the other stimuli respectively.

3.6.2 Latency to first approach categories

Placing latency to first approach data into categories can avoid the problems encountered when using censored data, such as possible under-representation of less responsive individuals in the data array. Separating individuals into four categories for each presentation (Table 3.5) enables the inclusion of all 68 animals' responses for

Table 3.4: ANOVA and post hoc comparisons (Sheffé) for differences in log transformed latency to first approach between stimuli. As for Table 3.1

Stimulus	MIR	TUB	MOU	EGG	RIT	RIW	POT	SYR
TUB	-0.099 1.000							
MOU	-0.209 0.240	-0.199 0.326						
EGG	-0.231 0.079	-0.221 0.126	0.0218 1.000					
RIT	0.1103 0.878	0.1202 0.827	**0.3192** **0.002**	**0.3410** **0.001**				
RIT	**0.3358** **0.001**	**0.3456** **0.001**	**0.5446** **0.001**	**0.5664** **0.001**	**0.2254** **0.050**			
POT	-0.046 1.000	-0.036 1.000	0.1627 0.642	0.1845 0.347	-0.156 0.523	**-0.381** **0.001**		
SYR	0.0151 1.000	0.0249 1.000	0.2239 0.169	**0.2458** **0.049**	-0.095 0.951	**-0.321** **0.001**	0.0613 0.998	
BAL	0.0573 0.998	0.0672 0.995	**0.2262** **0.038**	**0.2880** **0.006**	-0.053 0.951	**-0.278** **0.004**	0.1035 0.932	0.042 1.000

each stimulus (Figure 3.5). The categories do not cover equal time periods, but are organised so that there are enough occurrences in each category to allow statistical analysis, and non-occurrences can also be represented. The technique of visual presentation used in Figure 3.5 illustrates the proportion of animals not approaching for each stimulus (category 4). For all stimuli except for MOU and EGG, categories 1 and 2 contain the largest proportion of animals. This is the case even though the amount of time covered by these categories combined is the same as that for category 3. This supports the general impression that, in the majority of instances, individuals will approach a novel stimulus within one minute, if they are to approach at all. MOU and EGG, however, the stimuli with the most censored scores, also have the largest

Table 3.5: Latency time categories. Categories are not the same length of time, but characterise responses to allow inclusion of censored data (4) and statistical analyses (see Chapter Four).

Latency category	Latency to first occurrence
1	0-30
2	31-60
3	61-240
4	No occurrence within 240s

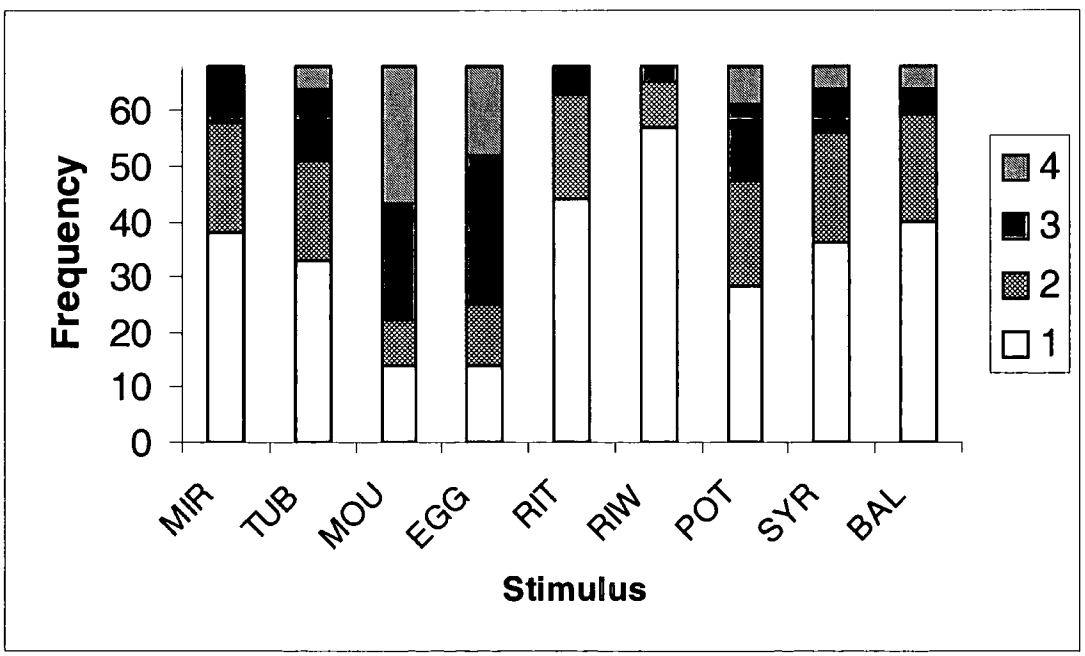


Figure 3.5: Bar chart of latency to first approach categories for nine stimuli. Categories are not time proportion equivalent (see Table 3.5). The x-axis denotes the stimulus, the y-axis cumulative frequency across the categories to a maximum of 68.

proportions for category 3, from one to four minutes, demonstrating that animals may hesitate yet still approach within the presentation time, in reaction to certain stimuli.

3.7 Latency to first contact

Latency to first contact was calculated in the same way as approach latency, but with time to physical contact with the stimulus measured rather than proximity to it. As with contact duration, first contact is defined as digital manipulation, or resting on the stimulus with hands or feet, but not brushing past or touching with the tail. Similarly to duration of proximity and duration of contact, latency scores are nested, with contact latency scores being lower by necessity than approach latency scores. It is not possible to touch a stimulus without first approaching it. Again, due to censored data in recording latency times, latency to first contact scores can be examined in two ways. Firstly, contact times were assessed with any censored scores excluded from analysis. Secondly, latency scores were calculated by converting times and censored data into categorical data (see table 3.5) to investigate the reactions of all animals to all stimuli. These categories were calculated in the same manner as for latency to first approach (section 3.5).

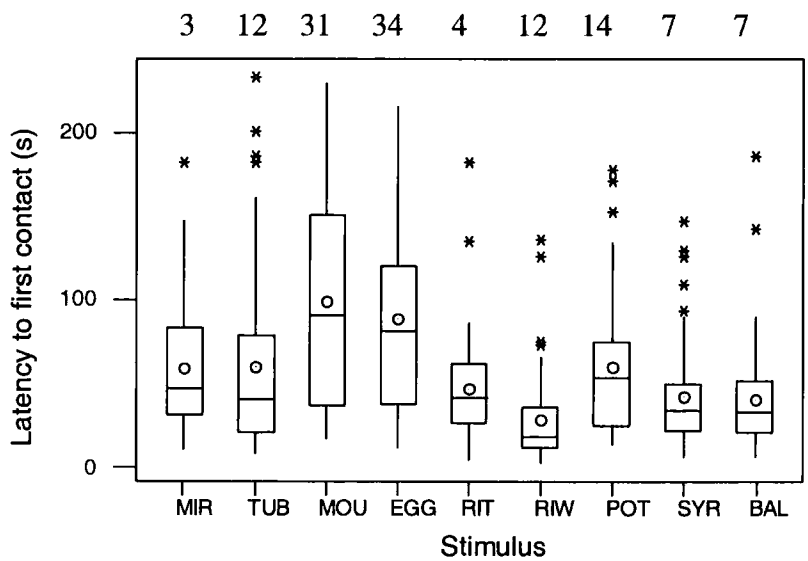


Figure 3.6: Box plot of latency to first contact with nine novel stimuli for 68 individuals. The x-axis shows stimulus code (see Table 2.2), the y-axis the latency to first contact time in seconds. Symbols as for figure 3.3

Table 3.6: ANOVA post hoc comparisons (Sheffé) for differences in log transformed latency to first contact between stimuli. As for Table 3.1.

Stimulus	MIR	TUB	MOU	EGG	RIT	RIW	POT	SYR
TUB	0.0557 0.998							
MOU	-0.186 0.374	-0.242 0.090						
EGG	-0.156 0.676	-0.210 0.266	0.0303 1.000					
RIT	0.0963 0.923	0.0406 1.000	**0.2825** **0.012**	0.2521 0.063				
RIT	**0.3871** **0.001**	**0.3314** **0.001**	**0.5733** **0.001**	**0.5429** **0.001**	**0.2908** **0.001**			
POT	0.0125 1.000	-0.043 1.000	0.1986 0.330	0.1683 0.618	-0.084 0.975	**-0.375** **0.001**		
SYR	0.1590 0.395	0.1032 0.914	**0.3451** **0.001**	**0.3148** **0.004**	0.0626 0.995	**-0.228** **0.043**	0.1465 0.591	
BAL	0.1652 0.336	0.1095 0.882	**0.3514** **0.001**	**0.3210** **0.003**	0.0689 0.991	-0.222 0.058	0.1528 0.529	0.0624 1.000

The data for means across individuals, when censored scores have been removed, have a mean of 55.71 and a standard deviation of 23.71 The maximum score of 116.86 indicates that no individual spent the majority of time in close contact with the presentation stimuli. These mean latencies to first contact data are normally distributed (K-S Z= 0.961, p=0.134).

3.7.1 Latency to close contact scores by stimulus

Examining the stimuli individually for latency times illustrates variation similar to that seen for latency to first approach (Figure 3.6). RIW has the lowest mean (28.25) and median (18.50) scores, but its IQR overlaps with that of RIT, SYR and BAL. Also, RIW has more censored values (12) than any of those other three. RIT no longer stands out as it did with the duration scores. MOU and EGG have means (98.6 and 89.21 respectively) and medians (91.0 and 81.5) higher than the Q3 (the top value of the IQR) of the seven other stimuli. Their IQRs are also much higher (MOU 113.5, EGG 82). A final factor separating them from the other stimuli is the high number of censored scores, with each of the two stimuli having almost half the individuals tested failing to come into contact with them. The lowest amount of censored data can be seen for MIR, the mirror. It can be seen from Figure 3.6 that there is a relationship between IQR, a

measure of spread of the data, and the number of censored scores (Spearman's correlation, r=0.826, p=0.006). This indicates that there is a relatively simple relationship between failing to come into contact with a stimulus and taking a long time to contact it within the presentation period, rather than some qualitative difference in response.

Several of the differences between stimuli indicated by Figure 3.6 are significant (see Table 3.6), when data are log transformed to give them a normal distribution. RIW has a significantly lower latency time to all stimuli except BAL. Note that this includes the other food stimulus, RIT, which is only significantly lower than MOU. SYR, conversely, is significantly lower scoring than three of the other stimuli. MOU and EGG are higher scoring than the lowest scoring four and three stimuli, respectively. This shows then that individuals' immediate reactions to the two food related stimuli are not the same; indeed, other, less apparently similar stimuli may be causing more similar first responses. As with latency to first approach, the least attractive stimuli are MOU and EGG.

3.7.2 Latency to first contact categories

The censored data excluded above can be introduced into analysis of latency to first contact if data are converted into categories (Figure 3.7). The larger proportions of

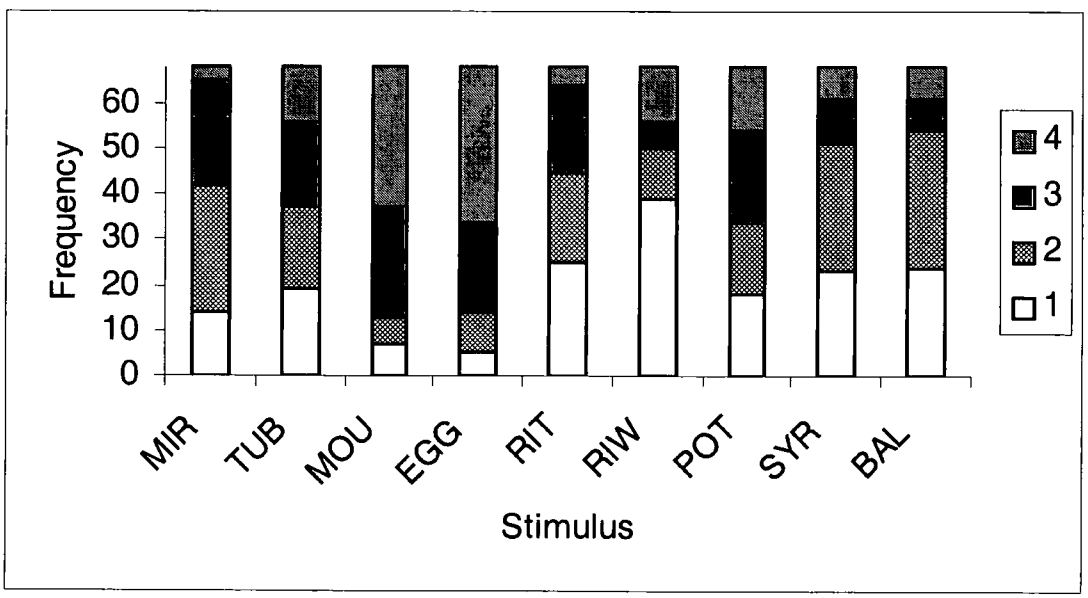


Figure 3.7: Bar chart of latency to first contact categories for nine stimuli. See Figure 3.5 and Table 3.5

animals failing to contact a stimulus than approach it can be seen if these results are compared to those in Figure 3.5. For EGG, half of the animals fall into category 4. MIR, SYR and BAL have more individuals in category 2 than category 1, indicating slightly more hesitation in contact than approach. All stimuli except MOU, EGG and POT, however, still contain the largest proportion of individuals in categories 1 and 2. Although it covers half of the possible presentation time, for no stimulus do half of the individuals fall into category 3.

3.8 Characterising stimulus response

Using the five measures described above, it is possible to characterise general responses to the nine stimuli, and assess their suitability for use in assessing responsiveness as a trait in the animals tested. In addition to this, certain measures may be more suitable for assessing specific stimuli, depending on aspects of the stimuli themselves and the characteristic marmoset response to them. Table 3.7 shows a reduced form of the above results (Sections 3.3 to 3.7). Stimuli are rated as low, average or high (relative to each other) for IQR and measures of central tendency across each of the 5 measures. Based on the above analysis, it is possible to separate the stimuli into four groups: (1) MIR, the mirror; (2) MOU and EGG, low duration, high latency stimuli; (3) RIT and RIW, the food-related stimuli; and (4) TUB, POT, SYR and BAL, simple novel stimuli. The four groups and the reasons for the groupings are described below.

3.8.1 The mirror

The mirror, MIR, shows a combination of responses that may be expected from its particular form. A relatively high visual attendance score shows that animals were paying attention to the stimulus. This is supported by the high proximity duration with a small spread, also indicating a standardised response. In contrast to this, neither latency to first contact nor contact duration are significantly above average. This array of scores fits the pattern that might be expected for a mirror, with animals approaching relatively quickly and with interest, but not touching the object as they interact with their own reflection and moving reflections of the room around them. Ongoing research of similar behaviours in baboons indicates heritability in variation in reaction to mirror presentation (Rogers *et al.*, 2002), especially with gaze aversion, which could be considered reciprocal to visual attendance. For this stimulus then, visual attendance and

Table 3.7: An overview of responses to nine stimuli using five measures of assessment. IQR refers to interquartile range, as an estimate of distribution of scores; MCT refers to measures of central tendencies, and is assessed on the relative position of both the mean and median. Scoring is relative. H= High (top three), A= Average (middle three), L= Low (lowest three). Where stimuli are tied on a border (i.e. joint third and fourth) an average score (A) is given. [1] Little variation between average (A) and low (L). [2] Latency scores with censored times removed from the analysis.

	Stimulus	**MIR**	**TUB**	**MOU**	**EGG**	**RIT**	**RIW**	**POT**	**SYR**	**BAL**
	Description	Mirror	Hollow grey plastic tube	Mouse and cheese dog toy	Cardboard egg tray	Raisin in transparent, sealed tube	Raisin in water filled bowl	Blue plastic pot	Transparent plastic syringe	Hollow pink plastic ball
Visual attendance	MCT	H	A	H	A	H	L	L	A	L
	IQR	L	A	A	H	L	A	L	H	A
Proximity duration	MCT	H	A	L	A	H	A	L	H	L
	IQR	L	A	H	H	L	L	A	H	A
Contact duration[1]	MCT	A	H	L	L	H	A	L	H	A
	IQR	A	H	L	L	H	L	A	H	A
Latency to first approach[1,2]	MCT	A	A	H	H	L	L	H	A	L
	IQR	L	H	H	H	A	L	A	L	A
Latency to first contact[1,2]	MCT	A	A	H	H	A	L	H	L	L
	IQR	A	H	H	H	A	L	A	L	L
Notes		Quick approach and inspection of reflections	Grouped with POT, SYR and BAL as simple "novel stimulus"	Animals attentive but refusing to touch. Due to "face" of stimulus?	Animals quite attentive but not touching	Animals spending a lot of time trying to retrieve food	Animals quick to approach and touch familiar bowl	Grouped with TUB, SYR and BAL as simple "novel stimulus	Grouped with TUB, POT and BAL as simple "novel stimulus	Grouped with TUB, POT and SYR as simple "novel stimulus

proximity duration are probably the most insightful measures to use in assessing response.

3.8.2 "Unattractive" stimuli

MOU, the mouse and cheese toy, and EGG, the cardboard egg tray, are two stimuli with similar responsiveness scores. The stimuli show the highest latency scores for both approach and contact latencies, with over 30 censored scores each for first contact. There are some differences in duration scores, as contact duration for MOU is very low, with an almost non-existent spread. EGG has the largest interquartile range for proximity duration, but both have similar measures of central tendency. Even where some variation in durations between the two stimuli does occur, differences are never significant. Both stimuli also have above average mean and medians for visual attendance. Animals are thus spending time looking at the stimuli, yet generally taking a very long time to approach and contact them or not doing so at all. A reaction such as this could be regarded as a negative response, in that the animal is aware of the stimulus, yet unwilling to interact closely with it. This style of response is easier to explain for MOU than EGG. The mouse head that is a prominent feature of MOU is approximately the same size of the head of a common marmoset. Individuals may have therefore been responding to this warily, treating MOU as an unusual social stimulus rather than merely a novel object. Although not measured objectively, it was noted that during MOU presentations many individuals were making threat or alarm vocalisations towards the stimulus, more so than for any other object used. Such a response to EGG is less expected. In fact, in the same establishment as the breeding colony, husbandry staff use cardboard egg trays as enrichment for experimental marmoset pairs. It is possible that the majority of animals are simply uninterested in the tray. In that case, however, a low visual attendance score would have been expected. It is due to this above average visual attention, but reluctance to approach or touch the stimuli, that MOU and EGG can be grouped together.

3.8.3 Food related stimuli

Responses to the two food-related stimuli show clear differences to the other objects present, but not always the same differences. RIT, the raisin in a sealed tube, has the highest measures of central tendency for visual attendance, proximity duration and

contact duration, indicating that not only were the animals most interested in this stimulus, but this interest involved touching or manipulating it. In fact, RIT is significantly different to all other stimuli in these measures, except MIR for visual attendance. This significant increase in interest for a foraging "task" has also been found for pair-housed female marmosets (Majolo *et al.*, 2003). Conversely, it is with latency time recording that RIW, the raisin in the water-filled food bowl, stands out. It has significantly lower times for latency to first approach than all other stimuli, and lower latency to contact scores than all but BAL. The 12 censored scores for contact latency, however, show that not all animals would touch it. Latency scores for RIT are not as distinctive, falling within a similar range to the non food-related stimuli. The variation in response between RIT and RIW can be explained in terms of the differences between the stimuli. RIT is a novel object that contains food that can be seen but not accessed, whereas RIW is a familiar food tray, with unfamiliar open water covering the food reward. The novel aspect of each is thus different. Animals are quickly approaching and touching a familiar food dish, but not spending a lot of time near it or touching it after finding the unfamiliar water blocking access to the raisin. Conversely, individuals take longer to approach, sniff at and touch the novel tube of RIT, but then spend a lot of time touching it, trying to recover the raisin, which is a familiar and preferred food. This reasoning would account for differences between the stimuli, but also explains why they can be grouped together. Although reactions to them are different, what makes responses to them vary from the remaining stimuli in both cases is the presence of food. The results here demonstrate that this presence elicits a qualitatively different response to other novel stimuli. Although not suitable as a simple novel stimulus due to these food-motivated differences in responsiveness, the stimuli are still useful in an investigation into behavioural variation in the marmoset. Individual, sex, and age related variation in response to food related tasks have previously been studied in callitrichids (e.g. Box *et al.*, 1999, Day *et al.*, 2003, Rogers, 1999; see Chapter One for discussion and more examples). With a sample size of 68 animals, data on food related stimuli from this study can add substantially to this area of research, as well as complementing responsiveness data gathered from other stimuli.

3.8.4 "Novel" stimuli

The fourth grouping of stimuli includes TUB, the grey plastic tube, POT, the blue pot, SYR, the syringe, and BAL, the pink hollow ball. The most useful description of this

group would simply be "novel stimuli". The only significant differences in response are seen between SYR and the others in duration times, due to animals spending more time near and touching the syringe than the other three objects. Due to the lack of unusual or distinctive response in this group, they make the best candidates for an unbiased assessment of novelty response in the marmosets studied, unaffected by additional factors. When selecting novel stimuli for enrichment purposes, both complexity and controllability are attractive attributes, with controllability being the more important (Sambrook & Buchanan-Smith, 1997). Whereas the complexity of these stimuli, as with any, would be difficult to ascertain, they can all be manipulated by the marmosets. All four can be moved by the animals, and have holes or crevices for them to explore. It is possible that the high latency times and low contact durations associated with EGG were because it did not appear easily moveable or manipulable. It is also possible that the actual size of EGG affected the animals' response, as the area of the shelf it covered was larger than any of the other stimuli. This large size may have affected the animals' willingness to interact.

3.9 Using the stimuli in further analysis

The possible confounding aspects of five of the stimuli (MIR, MOU, EGG, RIT and RIW) do necessarily not rule them out of any further analysis. It is sensible to avoid grouping all the stimuli together for a single assessment, but the variation seen in response is worth investigating in its own right. Analysis in the subsequent investigation will thus use TUB, POT, SYR and BAL as a novel stimulus selection for assessment of variation in responsiveness. Separate analysis of the remaining stimuli of interest, namely MIR, and the food stimuli RIT and RIW, will also be carried out. MOU is still a candidate for further analysis, because there is the possibility of reaction to it as a social stimulus that will highlight characteristic variation. Grouping objects can simplify analysis, but is also important to study the stimuli individually, in case there are any small but potentially important variations within groups. The next chapter focuses on the affect of sex, age and weight on responses to individual stimuli.

Chapter Four

The Effect of Age, Sex and Weight on Response

4.1 Introduction

When investigating individual variation in a trait, it is important to understand the different underlying factors that may cause that variation. The study of sex differences in behaviour may be affected by the age of the animals studied, and vice versa. Presence of confounding variables will affect the assessment of behavioural data for individual variation in the "temperament" of individuals. This chapter addresses this with relation to the sex of the animals studied, their age at testing, and their weight at testing. As stated in Chapter One (1.4), the three relevant hypotheses are:

2. Sex affects response to novel stimuli in common marmosets, with females being more responsive than males

2. Age affects response to novel stimuli in common marmosets, with older subadults being the most responsive

3. Weight affects response to novel stimuli in common marmosets

4.2 Analysis

To test the above hypotheses two methods of assessment were used. Each behavioural measure was assessed by testing individual stimulus scores, and also using a mean score across all stimuli. Stimuli were not split into the groups described in Chapter Three. This was to avoid a false reduction in the number of analyses due to a post-hoc decision on how to deal with the data.

Possible sex differences were thus investigated for each of the five measures, and for each stimulus. Where data were not suitable for parametric testing, Mann-Whitney tests were used. Otherwise, t-tests were carried out. Multiple comparisons were controlled for using a Bonferroni correction (Sokal & Rohlf, 1994). This sets the p-value needed to

accept a significant difference between groups at 0.055. The Bonferroni method is a conservative method for controlling for multiple comparisons during statistical testing. Using a Bonferroni correction method, each comparison uses a significance level of $\alpha'' = \alpha/k$, where α= the chosen level of significance testing (i.e. 0.05), α''= the level of testing after adjustment and k = the number of comparisons. In this instance then, $\alpha'' = 0.05/9 = 0.0055$, as for each behavioural measure on each variable, nine stimuli are being evaluated. Where mean scores across stimuli are tested, the required p-value is 0.01.

The effect of age and weight were dealt with differently, as they are continuous rather than dichotomous variables. For each stimulus within each measurement, a regression analysis was used to investigate the significance and size of any relationship between the demographic and the response measures. Due to the age of the animals, their weight may be affected considerably by age. In the analysis, therefore, age was controlled for when assessing the effect of weight on response. To control for multiple comparisons, a Bonferroni correction factor was used. Where measurement data were not continuous, but in ordinal categories, (i.e. latency category data), ordinal logistic regressions were used. Again, multiple comparisons were controlled for with Bonferroni corrections.

4.2.1 Power

Most statistical analyses concentrate on minimizing or controlling the possibility of a Type I error (α). There is of course always the possibility of a Type II error (β), failing to find a difference that is actually there. Power analysis deals with this problem, power being $1-\beta$. A more powerful experiment then is one that is more likely to reject a false H_0 than a less powerful one (Howell, 1997). Cohen (1992) suggests a power of 0.80 as a convention for general use. This level is a balance between incurring too great a risk of a Type II error and having a demand for a sample size (N) that is "likely to exceed the investigator's resources" (Cohen, 1992). Power analysis was carried out for all parametric tests.

Non-significant results are likely to be an important part of these analyses, if confounding variables are to be dismissed. Bonferroni corrections for multiple comparisons make non-significant results more likely to occur, as they lower required p-values. Due to this, it would be helpful to demonstrate that those results were actually

because of there being no difference, rather than because the statistical test did not pick those differences up. An a priori power analysis shows that with an alpha of 0.05 (i.e. 0.05 being the required level of significance for acceptance of differences), a power of 0.80 would require a sample size of 128 individuals. During testing, due to time constraints, only just over half that number (N=68) were tested. The best path to take from here was to analyse the data that were collected within the time constraints and report the power of the tests using a post-hoc power analysis for all non-significant results. All such analyses were carried out in GPower (Buchner *et al.*, 1993; Erdfelder *et al.*, 1996).

The power of tests was calculated in two ways. Firstly, effect sizes were calculated from the data themselves in GPower. They were then used to calculate the power of the test. Secondly, effect size conventions (quoted in Gpower, Buchner *et al.*, 1993; Erdfelder, *et al.*, 1996) for a medium effect were also used to calculate the power. These two methods are referred to as calculated power and convention power throughout the chapter.

4.3 The effect of sex on responsiveness

4.3.1 Visual attendance

No significant differences were found between males and females in their mean attendance scores (two sample t-test, T_{60}=-2.23,p=0.030), when corrections for multiple comparisons are taken into account. When data are examined stimulus by stimulus (Table 4.1), only one, TUB, shows differences between the sexes after a Bonferroni correction for multiple comparisons ($t_{(61)}$= -3.36, p= 0.0014). Although no sex differences were found for eight of the stimuli, the calculated power of the tests are generally low. When effect size is calculated for each test only two stimuli, MOU and BAL having a power approaching or above 0.8 (0.74 and 0.85 respectively). The remaining tests have powers of below 0.6. If, however, Cohen's effect size convention of 0.5 for a medium effect is used, power is raised so that five stimuli score above 0.7 (MIR, MOU, RIW, SYR, BAL). The power for RIT remains at around 0.3 for both calculations. This stimulus would show significant sex differences without the correction for multiple comparisons.

Table 4.1: Sex differences in visual attendance for nine stimuli, two-sample t-test and power analysis results. Calculated power show the power of the test using an effect size calculated from sample means and population standard deviation. Convention power used Cohen's (1977) convention effect size of 0.5. Stimulus codes are from Table 2.2.

Stimulus	Test Statistic	p-value	Effect size	Calculated power	Convention power
MIR	$t_{(53)}$= -0.77	0.44	0.1922	0.5602	0.8929
TUB	$t_{(61)}$= -3.36	0.0014			
MOU	$t_{(62)}$= -0.34	0.73	0.0814	0.7434	0.9607
EGG	$t_{(45)}$= -1.59	0.12	0.4121	0.5332	0.6668
RIT	$t_{(56)}$= -2.49	0.016	0.5045	0.3330	0.3262
RIW	$t_{(56)}$= -1.24	0.22	0.3027	0.4992	0.7787
POT	$t_{(56)}$= -2.09	0.041	0.4993	0.4704	0.4715
SYR	$t_{(56)}$= -1.35	0.18	0.3358	0.5020	0.7423
BAL	$t_{(59)}$= -0.03	0.98	0.0062	0.9800	0.9973

4.3.2 Proximity duration

There were no differences between sexes in the amount of time spent in proximity to the stimuli (Two sample t-test, T_{59}=-1.72, p=0.092). Stimuli were assessed individually for differences using Mann-Whitney tests with a Bonferroni correction (Table 4.2). There are no significant sex differences for any of the nine stimuli, although RIT shows a difference that would be significant without correcting for multiple comparisons.

Table 4.2: Sex differences in proximity duration for 68 animals across nine stimuli, using Mann-Whitney tests. Stimulus codes are from Table 2.2.

Stimulus	Test Statistic	p-value
MIR	U= 458.0	0.267
TUB	U= 455.0	0.251
MOU	U= 510.5	0.644
EGG	U= 483.0	0.424
RIT	U= 344.5	0.011
RIW	U= 432.0	0.150
POT	U= 470.0	0.337
SYR	U= 445.5	0.205
BAL	U= 537.0	0.910

Table 4.3: Sex differences in contact duration for 68 animals across nine stimuli, using Mann-Whitney tests. Stimulus codes are from Table 2.2.

Stimulus	Test Statistic	p-value
MIR	U= 545.0	0.990
TUB	U= 498.0	0.540
MOU	U= 521.0	0.718
EGG	U= 446.0	0.171
RIT	U= 427.5	0.135
RIW	U= 394.0	0.055
POT	U= 431.0	0.137
SYR	U= 439.5	0.178
BAL	U= 500.0	0.556

4.3.3 Contact duration

There were no significant differences between sexes either for the log transformed mean duration of contact per individual (two sample t-test, t_{54}=-1.48, p=0.15), or when data for the stimuli were examined individually (Table 4.3).

4.3.4 Latency to first approach

Mean scores for each animal for latency to approach, log transformed with censored scores removed from the analysis, show no significant differences between the sexes (two-sample t-test, $t_{(47)}$= 1.48, p= 0.14). This is also the case when stimuli are tested individually (Table 4.4). Only two stimulus tests have a power above 0.8, TUB (0.87) and MOU (0.81) when power is calculated from effect sizes based on the data. The remaining t-tests have a power of between 0.48 and 0.62. If effect size conventions are used, five of the stimuli have a power of above 0.8 (TUB, MOU, EGG, POT, BAL), but RIW's power is lowered to 0.28. When censored values are included in the analysis, and data are assessed in categories (see section 3.52), SYR does show a significant difference between the sexes (Mann-Whitney, U=320, p=0.004), with females taking longer to approach than males (Table 4.5).

4.3.5 Latency to first contact

Mean contact latency times with censored data removed show no differences due to sex ($t_{(53)}$= 1.51, p= 0.14). When tested individually (Table 4.6), only one stimulus, RIT,

Table 4.4: Sex differences in latency to first approach for 68 animals across nine stimuli. For power, see table 4.1. Censored time data were removed from analysis; sample sizes for each test are given. All tests are regression analyses

Stimulus	N (F/M)	Test Statistic	p-value	Effect size	Calculated power	Convention power
MIR	42/26	$t_{(57)}= 2.01$	0.049	0.4943	0.4933	0.5023
TUB	40/24	$t_{(33)}= 0.16$	0.87	0.0887	0.8744	0.9797
MOU	26/17	$T_{(31)}= 0.24$	0.81	0.0771	0.8156	0.9461
EGG	34/18	$T_{(34)}= 0.68$	0.50	0.2259	0.6132	0.8586
RIT	42/26	$T_{(55)}= 1.66$	0.10	0.4079	0.4898	0.6325
RIW	42/26	$T_{(59)}= 2.62$	0.011	0.6437	0.5940	0.2783
POT	36/25	$T_{(51)}= 1.01$	0.32	0.2453	0.5042	0.8225
SYR	38/26	$T_{(40)}= 1.70$	0.097	0.4474	0.5137	0.6116
BAL	40/24	$T_{(52)}= 0.99$	0.32	0.2513	0.5148	0.8266

Table 4.5: Sex differences in latency to first approach category data for nine stimuli, using Mann-Whitney tests. Stimulus codes are from Table 2.2.

Stimulus	Test Statistic	p-value
MIR	U = 402	0.069
TUB	U = 528.5	0.825
MOU	U = 513	0.669
EGG	U = 526	0.804
RIT	U = 426	0.131
RIW	U = 340.5	0.009
POT	U = 432.5	0.152
SYR	U =320.5	0.004
BAL	U = 482.5	0.423

demonstrates any differences between the sexes, with females taking longer to touch the stimulus than males ($t(45)= 3.04$, $p= 0.0039$). The other food related stimulus, RIW does tend towards significance ($t(45)= 2.23$, $p=0.030$), but is not because of the Bonferroni correction factor being used. The eight stimulus tests that show no significant variation have a relatively low power, all between 0.5 and 0.64, when calculated effect sizes are used. Convention effect sizes (Cohen, 1992) give a power of near or above 0.8 to MIR, TUB, MOU, POT and BAL, but the power for the RIT test becomes very low. If censored data are included in a non-parametric analysis (see section 3.6.2), both RIT (Mann-Whitney, $U= 285$, $p= 0.001$) and RIW (Mann-Whitney, $U= 264$, $p<0.001$) show males coming into contact more quickly than females (Table 4.7).

Table 4.6: Sex differences in latency to first contact for 68 animals across nine stimuli. For power see table 4.1. Censored time data were removed from analysis; sample sizes for each test are given. All tests are regression analyses.

Stimulus	N (F/M)	Test Statistic	p-value	Effect size	Calculated power	Convention power
MIR	40/25	$T_{(38)}= 0.59$	0.56	0.1598	0.6306	0.9211
TUB	34/22	$T_{(35)}= 0.76$	0.45	0.2168	0.5750	0.8620
MOU	23/14	$T_{(24)}= 0.72$	0.48	0.2475	0.5842	0.7922
EGG	18/16	$T_{(29)}= 1.13$	0.27	0.3929	0.5245	0.6379
RIT	38/26	$T_{(45)}= 3.04$	0.0039	0.7983		
RIW	30/26	$T_{(45)}= 2.23$	0.030	0.5947	0.5004	0.3652
POT	30/24	$T_{(49)}= 0.91$	0.37	0.2495	0.5400	0.8251
SYR	36/25	$T_{(40)}= 1.69$	0.098	0.4618	0.5396	0.5964
BAL	37/24	$T_{(58)}= 0.54$	0.59	0.1394	0.5396	0.9214

Table 4.7: Sex differences in latency to first approach category data for nine stimuli, using Mann-Whitney tests. Stimulus codes are from Table 2.2.

Stimulus	Test Statistic	p-value
MIR	U = 497.5	0.540
TUB	U = 473	0.355
MOU	U = 531	0.842
EGG	U = 402.5	0.053
RIT	U = 285	0.001
RIW	U = 264	0.000
POT	U = 368	0.024
SYR	U =358.5	0.018
BAL	U= 498	0.544

4.3.6 Overall effects of sex on responsiveness

Mean scores for the five measures show no relationship with the sex of the animals. Across all individual stimulus analyses only four stimuli show any significant variation. Males visually attend to TUB for longer, and have lower latency to close contact times than females for RIT, a food related stimulus. Analysis of category data shows an additional significant sex difference for latency to approach SYR. Also, latency to first contact demonstrates differences for both of the food stimuli, rather than just RIT. In all three of these cases, males are more responsive than females.

Overall, the lack of sex differences refutes Hypothesis 2, which states that sex differences should be found, with females being more responsive than males. Marmosets might be considered a special case when it comes to sex differences in behaviour. Unlike many primate (or indeed mammal) species, males take much of the parental care responsibilities. This would lead to less of a reproductive skew than might be expected from Bateman's principle (Bateman, 1948; Futuyma, 1998, p587), and so marmoset males may not be in general more responsive or risk taking than females. The fact that more differences, and non significant tendencies toward differences, where males are more responsive were seen for the food related stimuli than any others is the reverse finding than what would be predicted from previous experimental studies where females are more responsive (e.g. Box, 1988, Visalberghi *et al.*, 2003). This finding does follow the pattern found for innovative (rather than novelty responsive) behaviours across all primates (Reader & Laland, 2001), which included in the main behaviours related to novelty. It is possible that sex differences seen in family or pair groups is due to female primacy of access to foods rather than a quicker responsiveness *per se*. In this case, the quicker responses of males may be because of the individual animal test set up, or that animals are housed in same sex peer groups. Looking at mixed sex housing, even within the same colony as testing took place, might find the differences reversed if females are present to affect male performance.

Across all measures relatively large standard deviations, indicating variability in responses, caused analyses to have relatively low power when effect sizes were calculated from the data. However, if Cohen's (1992) suggested convention effect sizes are used in the power calculations, the power of tests are generally above 0.8. The varying sample sizes for individual stimulus latency scores also have an effect on the power of calculations. Sample size from the study was constrained by time. Also, it could not be predicted beforehand how many animals would fail to approach or touch the various stimuli, leading to censored scores in the latency data. The generally higher sample sizes for food related stimuli (due to fewer animals having censored scores), as well as greater differences between males and females in scores compared to standard deviations all affect the relative power of the tests. Power for RIT and RIW is generally lower than for the other stimuli across all measures.

4.4 The effect of age on responsiveness

The age of the animals tested is normally distributed (K-S Z= 1.250, p= 0.088), allowing parametric statistics to be used in analysis where the data being tested are also normally distributed.

4.4.1 Age and visual attendance

Regression analysis demonstrates no significant negative relationship between age at testing and mean visual attendance when the required p-value is corrected for multiple comparisons (r^2=0.070, $F_{(1,66)}$= 4.939,p= 0.030), although there is a non-significant indication of a very small effect. When each stimulus is considered individually, only one, TUB, shows any significant relationship (R^2= 0.143, $F_{(1,66)}$= 10.972, p=0.002). SYR (R^2= 0.068, $F_{(1,66)}$= 4.505, p= 0.038) would have shown a significant relationship if corrections for multiple comparisons had not been made. Using calculated power, only three of the stimuli, MIR, RIT and RIW, have a power of above 0.8 (although POT is close). When a standard medium effect size is used, convention power is above 0.85 for all of the stimuli that show no significant relationship.

Table 4.8: Relationship between age and visual attendance for 68 animals in response to nine novel stimuli, using regression analysis. Power is calculated in two ways, with an effect size calculated from the data, and using Cohen's effect size convention for a medium effect, 0.15 (Cohen, 1988)

Stimulus	Test Statistic (Regression)	p-value	Effect size	Calculated power	Convention power
MIR	R^2= 0.001, $F_{(1,66)}$= 0.063	0.802	0.0010	0.8085	0.9987
TUB	R^2= 0.143, $F_{(1,66)}$= 10.972	0.002			
MOU	R^2= 0.026, $F_{(1,66)}$= 1.737	0.192	0.0266	0.5167	0.9691
EGG	R^2= 0.007, $F_{(1,66)}$= 0.456	0.502	0.0070	0.5938	0.9941
RIT	R^2= 0.000, $F_{(1,66)}$= 0.003	0.958	0	0.9580	1.0000
RIW	R^2= 0.000, $F_{(1,66)}$= 0.032	0.858	0	0.8580	1.0000
POT	R^2= 0.001, $F_{(1,66)}$= 0.087	0.769	0.0010	0.7765	0.9984
SYR	R^2= 0.068, $F_{(1,66)}$= 4.505	0.038	0.0729	0.5459	0.8596
BAL	R^2= 0.005, $F_{(1,66)}$= 0.304	0.583	0.0050	0.6423	0.9960

4.4.2 Age and proximity duration

A regression analysis of the effect of age on the log mean scores of all individuals fails to show a significant relationship at a 0.05 level (r^2=0.042, $F_{(1,66)}$=2.90, p=0.093). Individual stimulus data are not normally distributed, so category data derived from the

Table 4.9: Categories for non-normally distributed duration time analysis. Categories are not time proportion equivalent, but characterise responses to allow statistical analysis.

Duration category	Duration time (s)
1	0-30
2	31-91
3	91-240

Table 4.10: Relationship between age and proximity duration for 68 animals in response to nine novel stimuli, using ordinal logistic regression analysis.

Stimulus	Test Statistic (logistic regression)	p-value
MIR	N= 68, Z=1.33	0.183
TUB	N= 68, Z= 1.45	0.147
MOU	N= 68, Z= 0.91	0.362
EGG	N= 68, Z= -0.41	0.681
RIT	N= 68, Z= 0.38	0.701
RIW	N= 68, Z= 0.59	0.558
POT	N= 68, Z= 0.95	0.341
SYR	N= 68, Z= 1.72	0.086
BAL	N= 68, Z= 1.38	0.167

Table 4.11: Relationship between age and contact duration for 68 animals in response to nine novel stimuli, using ordinal logistic regression analysis.

Stimulus	Test Statistic (logistic regression)	p-value
MIR	N= 68, Z= 0.25	0.804
TUB	N= 68, Z= 0.67	0.504
MOU	N= 68, Z= 1.50	0.134
EGG	N= 68, Z= 0.82	0.415
RIT	N= 68, Z= 1.36	0.174
RIW	N= 68, Z= 0.70	0.482
POT	N= 68, Z= 1.51	0.132
SYR	N= 68, Z= 1.41	0.160
BAL	N= 68, Z= 0.67	0.501

scores must be used in ordinal logistic regressions (Table 4.9). When category data are assessed stimulus by stimulus, none of the tests show a significant relationship between age and proximity duration (table 4.10).

4.4.3 Age and contact duration

Mean duration of contact scores show no significant relationship with age when multiple comparisons are controlled for (R^2= 0.068, $F_{(1,66)}$= 4.54, p=0.037), although there are non-significant indications of a possible relationship. As with proximity duration, individual stimulus data are not normally distributed. Therefore, in analysis, response categories are used (Table 4.9). Again, no significant differences are seen, when data are analysed in this way.

Table 4.12: Relationship between age and mean latency to first approach times, with censored data removed, across nine stimuli. Sample size changes due to censored scores. See Table 4.8.

Stim.	N	Test Statistic (Regression)	p-value	Effect size	Calculated power	Convention power
MIR	68	R^2= 0.032, $F_{(1,66)}$= 2.21	0.142	0.0330	0.5085	0.9553
TUB	64	R^2= 0.016, $F_{(1,66)}$= 1.01	0.320	0.0162	0.5301	0.9818
MOU	43	R^2= 0.004, $F_{(1,35)}$= 0.14	0.712	0.0040	0.7347	0.9868
EGG	54	R^2= 0.015, $F_{(1,50)}$= 0.74	0.395	0.0152	0.5606	0.9766
RIT	68	R^2= 0.055, $F_{(1,66)}$= 3.87	0.053	0.0582	0.5105	0.8875
RIW	68	R^2= 0.044, $F_{(1,66)}$= 3.03	0.086	0.0460	0.5130	0.9252
POT	61	R^2= 0.000, $F_{(1,59)}$= 0.02	0.903	0.0000	0.9030	0.9990
SYR	64	R^2= 0.000, $F_{(1,62)}$= 0.00	0.975	0.0000	0.9750	1.0000
BAL	64	R^2= 0.022, $F_{(1,62)}$= 1.38	0.245	0.0224	0.5206	0.9725

Table 4.13: Relationship between age and latency to first approach categories for 68 animals across nine stimuli. Ordinal logistic regression analyses are used. For categories see Table 3.5. N= 68

Stimulus	Test Statistic (logistic regression)	p-value
MIR	Z = -0.43	0.666
TUB	Z = 1.30	0.194
MOU	Z = -0.89	0.373
EGG	Z = -0.30	0.760
RIT	Z = -2.12	0.034
RIW	Z = -1.75	0.080
POT	Z = 0.77	0.444
SYR	Z = 0.10	0.919
BAL	Z= -1.34	0.180

4.4.4 Age and latency to first approach

Latency to first approach shows no significant relationship with the animals' age at testing, using a log transformed mean measure (r2= 0.017, F(1,66)= 1.11, p= 0.295). Analyzing stimuli individually, using log-transformed times with censored data removed, there are also no significant relationships with age. Analyzing latency to first approach as category data including censored times (see Table 3.5) shows only one relationship, with older animals taking longer to approach RIT, at a level below 0.05 (logistic regression analysis, N=68, Z= -2.12, p=0.034) (Table 4.12).

Table 4.14: Relationship between mean latency to first contact times, with censored data removed, across nine stimuli. Sample size changes due to censored scores. See Table 4.8

Stim.	N	Test Statistic (Regression)	p-value	Effect size	Calculated power	Convention power
MIR	65	R^2= 0.016, $F_{(1,63)}$= 1.01	0.319	0.0162	0.5320	0.9827
TUB	56	R^2= 0.003, $F_{(1,54)}$= 0.14	0.708	0.0030	0.7305	0.9947
MOU	37	R^2= 0.002, $F_{(1,36)}$= 0.06	0.807	0.0020	0.8139	0.9873
EGG	34	R^2= 0.026, $F_{(1,32)}$= 0.87	0.359	0.0266	0.5422	0.9084
RIT	62	R^2= 0.078, $F_{(1,62)}$= 5.27	0.025	0.0845	0.5147	0.7716
RIW	56	R^2= 0.046, $F_{(1,54)}$= 2.59	0.114	0.0482	0.5180	0.9004
POT	54	R^2= 0.003, $F_{(1,52)}$= 0.14	0.710	0.0030	0.7316	0.9939
SYR	61	R^2= 0.005, $F_{(1,59)}$= 0.30	0.585	0.0050	0.6384	0.9935
BAL	61	R^2= 0.008, $F_{(1,59)}$= 0.49	0.486	0.008	0.5819	0.9900

Table 4.15: Relationship between age and latency to first contact categories for 68 animals across nine stimuli. Ordinal logistic regression analyses are used. For categories see table 3.5. N= 68.

Stimulus	Test Statistic (logistic regression)	p-value
MIR	Z = -1.80	0.070
TUB	Z = 0.50	0.620
MOU	Z = -1.91	0.057
EGG	Z = 0.60	0.546
RIT	Z = -1.48	0.139
RIW	Z = -0.20	0.843
POT	Z = 1.61	0.108
SYR	Z = 0.34	0.737
BAL	Z = -0.141	0.160

4.4.5 Age and latency to first contact

Mean latency to first contact scores show no relationship with age at testing (r^2= 0.041, $F_{(1,66)}$= 0.03, p<0.866). This is also the case when stimuli are tested individually, with only one stimulus, RIT, showing a relationship at less than p= 0.05, which still above the corrected significance of p=0.0055 (r^2= 0.078, $F_{(1,62)}$= 5.27, p=0.025). When categorical data including censored values (Table 3.5) are used to individually assess each stimulus (with ordinal logistic regression analysis), no significant relationships with age are seen (table 4.13).

4.4.6 Overall effects of age on responsiveness

Using mean scores to assess variation, none of the measures show significant variation with age, although visual attendance and contact duration both have alpha below 0.05, indicating that younger animals may be spending more time both looking at and in contact with novel stimuli. These possible patterns do not hold when individual stimuli are tested. Throughout all tests, only one stimulus shows significant variation; younger animals are more visually attentive to TUB. Younger animals can thus only said to be more responsive to this one stimulus, and the general lack of significant variation does not support Hypothesis 3, that age may have an effect on novelty responsiveness. This contrasts with findings from previous studies of mixed sex and family groups with lower sample sizes in several different primate species (white-fronted capuchins, Visalberghi & Fragaszy, 1995; Fragaszy *et* al., 1997; Visalberghi *et* al., 2003; baboons, Joubert & Vauclair, 1986; saddleback tamarins, Menzel & Menzel, 1979). These studies indicated that an effect of age existed for both novel and novel food related objects. It is possible, however, that differences in the current study may have been demonstrated if a greater age range were studied, as previous analyses include fully adult animals.

As with the analysis of sex differences, the low calculated power of the individual tests does not allow an unequivocal statement of the lack of any effect. In situations where power is above 0.8 (Cohen's, 1992, convention for accepting an non-significant result without worry of a Type II error), calculated effect sizes are effectively zero. This means that only when there are no registerable relationships at all, will a test have power. This changes radically if effect size conventions are used rather than effect sizes calculations from the data. In this case, the power for almost every stimulus test is

above 0.8, meaning that the lack of relationships seen are not Type II errors, but based on a real lack of relationship in the data.

4.5 Is there a relationship between weight at testing and responsiveness?

The animals tested span an age range where a significant amount of physical growth is still occurring. The weight of the animal may thus be related to its age. To control for this potential problem, the residuals of a regression of weight on age (r^2= 0.403, F= 44.502, p< 0.001) were used in the following analysis. Both weight (K-S Z= 0.895, p= 0.400) and the residuals of weight on age (K-S Z= 0.745, p= 0.635) are normally distributed.

4.5.1 Visual attendance

There is no significant relationship between weight at testing and mean visual attendance (r^2= 0.03, F= 0.197, p=0.658). When stimuli are tested individually, no significant relationships are seen (Table 4.16), although SYR has a p-value of below 0.05 (r^2= 0.064, F= 4.505, p=0.038). When power is calculated with effect sizes from the data, five stimulus tests have a power approaching or above 0.8 (MIR, 0.8085; RIT, 0.9580; RIW, 0.8580; POT, 0.7765; SYR, 0.7765). When power is calculated using Cohen's convention for a medium effect, all stimuli have a power above 0.8.

Figure 4.16: Regression of weight on visual attendance for 68 animals across nine stimuli. See Table 4.8.

Stimulus	Test Statistic (Regression)	p-value	Effect size	Calculated power	Convention power
MIR	R^2= 0.001, F= 0.63	0.802	0.001	0.8085	0.9978
TUB	R^2= 0.013, F= 0.898	0.347	0.0131	0.5301	0.9876
MOU	R^2= 0.026, F= 1.737	0.192	0.0266	0.5167	0.9691
EGG	R^2= 0.007, F= 0.456	0.502	0.0070	0.5938	0.9941
RIT	R^2= 0.000, F= 0.000	0.958	0.000	0.9580	1.0000
RIW	R^2= 0.000, F= 0.032	0.858	0.000	0.8580	1.0000
POT	R^2= 0.001, F= 0.087	0.769	0.010	0.7765	0.9986
SYR	R^2= 0.064, F= 4.505	0.038	0.0683	0.7765	0.8569
BAL	R^2= 0.005, F= 0.304	0.583	0.0050	0.6423	0.9960

4.5.2 Proximity duration

The mean proximity duration of the 68 animals shows no significant relationship with their weight (r^2= 0.006, F= 0.367, p= 0.547). For individual assessment of stimuli, category data must be used because of the non-normal distribution of proximity data (Table 4.9). No significant relationships are seen (Table 4.17), although two stimuli, EGG and SYR, have p-values of less than 0.05.

4.5.3 Contact duration

Mean contact duration times show no relationship with weight at testing (r^2= 0.005, F= 0.331, p= 0.567). The non-normally distributed data means that categories (Table 4.9)

Table 4.17: Ordinal logistic regression of weight on proximity duration for 68 animals across nine stimuli.

Stimulus	Test Statistic (logistic regression)	p-value
MIR	Z= -0.14	0.890
TUB	Z= -0.80	0.421
MOU	Z= -1.34	0.179
EGG	Z= -2.04	0.041
RIT	Z= 0.88	0.379
RIW	Z= -0.07	0.945
POT	Z= -1.21	0.227
SYR	Z= -2.33	0.020
BAL	Z= -1.36	0.175

Table 4.18: Ordinal logistic regression of weight on contact duration for 68 animals across nine stimuli.

Stimulus	Test Statistic (logistic regression)	p-value
MIR	Z= 1.59	0.112
TUB	Z= -1.08	0.280
MOU	Z= -1.42	0.157
EGG	Z= -1.32	0.188
RIT	Z= 0.90	0.369
RIW	Z= -0.24	0.812
POT	Z= -0.97	0.330
SYR	Z= -2.18	0.029
BAL	Z= -1.88	0.060

and logistic regressions are used for individual stimulus analysis. As for mean times, no stimuli when tested individually show any significant relationship.

4.5.4 Latency to first approach

When censored values are removed from the analysis, mean log transformed latency to first approach scores show no relationship with weight (r^2= 0.044, F= 3.050, p=0.085). When the stimuli are tested individually, again using time scores and removing censored data, no significant relationships are seen (Table 4.19). Calculated power analysis of the tests shows four, TUB, MOU, EGG and RIW to have a power approaching or above 0.8. If power is calculated using effect size conventions, all stimulus tests have a power above 0.8. If censored values are included in the analysis

Table 4.19: Relationship between mean latency to first approach times and weight, with censored data removed, across nine stimuli. See table 4.8.

Stim.	N	Test Statistic (Regression)	p-value	Effect size	Calculated power	Convention power
MIR	68	R^2= 0.014, $F_{(1,66)}$= 0.96	0.331	0.0141	0.5271	0.9865
TUB	64	R^2= 0.000, $F_{(1,62)}$= 0.01	0.943	0.0000	0.9430	1.0000
MOU	43	R^2= 0.003, $F_{(1,41)}$= 0.14	0.709	0.0030	0.7263	0.9866
EGG	54	R^2= 0.003, $F_{(1,50)}$= 0.14	0.715	0.0030	0.7362	0.9941
RIT	68	R^2= 0.046, $F_{(1,66)}$= 3.18	0.079	0.0482	0.5132	0.9192
RIW	68	R^2= 0.000, $F_{(1,66)}$= 0.03	0.868	0.0000	0.8680	1.0000
POT	61	R^2= 0.015, $F_{(1,59)}$= 0.92	0.342	0.0152	0.5317	0.9805
SYR	64	R^2= 0.006, $F_{(1,62)}$= 0.40	0.531	0.0060	0.6031	0.9933
BAL	64	R^2= 0.026, $F_{(1,62)}$= 1.67	0.201	0.0266	0.5120	0.9640

Table 4.20: Relationship between weight at testing and latency to first approach categories for 68 animals across nine stimuli.

Stimulus	Test Statistic (logistic regression)	p-value
MIR	Z= 1.44	0.255
TUB	Z= 1.93	0.053
MOU	Z= 1.24	0.215
EGG	Z= 1.61	0.106
RIT	Z= -0.22	0.827
RIW	Z= -1.03	0.303
POT	Z= 1.68	0.062
SYR	Z= 0.23	0.818
BAL	Z= -0.17	0.861

and the data put into categories (Table 4.9), no significant relationships with age are seen (Table 4.20).

4.5.5 Latency to first contact

There is no significant relationship between animals' weight at testing and mean latency to first contact with censored data removed from the analysis (r^2= 0.041, F= 2.818, p= 0.098) When examined individually, removing censored data and using a log transformation of times, none of the stimuli show significant relationships (Table 4.21). Four of the stimulus tests have a calculated power of above 0.8 (MIR, EGG, SYR and BAL), the remainder being between 0.5 and 0.6. If the power analysis is based on convention effect sizes, all stimuli have a power of above 0.8. When censored values

Table 4.21: Relationship between mean latency to first contact times and weight, with censored data removed, across nine stimuli. See table 4.8.

Stim.	N	Test Statistic (Regression)	p-value	Effect size	Calculated power	Convention power
MIR	65	R^2= 0.001, $F_{(1,63)}$= 0.04	0.835	0.0010	0.8402	0.9986
TUB	56	R^2= 0.053, $F_{(1,54)}$= 3.02	0.088	0.559	0.5160	0.8755
MOU	37	R^2= 0.026, $F_{(1,35)}$= 0.92	0.344	0.0266	0.5332	0.9157
EGG	34	R^2= 0.002, $F_{(1,32)}$= 0.05	0.822	0.0020	0.8278	0.9855
RIT	62	R^2= 0.011, $F_{(1,62)}$= 0.71	0.403	0.0111	0.5445	0.9863
RIW	56	R^2= 0.010, $F_{(1,54)}$= 0.53	0.470	0.0101	0.5815	0.9851
POT	54	R^2= 0.060, $F_{(1,52)}$= 3.33	0.074	0.0638	0.5165	0.8450
SYR	61	R^2= 0.001, $F_{(1,59)}$= 0.06	0.803	0.0010	0.8088	0.9978
BAL	61	R^2= 0.000, $F_{(1,59)}$= 0.00	0.976	0.0000	0.9760	1.0000

Table 4.22: Relationship between weight at testing and latency to first contact categories for 68 animals across nine stimuli. Tests are logistic regressions.

Stimulus	Test Statistic (logistic regression)	p-value
MIR	Z = -1.60	0.110
TUB	Z = 1.28	0.199
MOU	Z = 0.00	0.999
EGG	Z = 2.24	0.025
RIT	Z = -1.17	0.243
RIW	Z = -0.90	0.370
POT	Z = 1.60	0.110
SYR	Z = 0.58	0.562
BAL	Z = -0.14	0.892

are included in the analysis and the data are grouped in categories (see table 3.5), no significant relationships are seen with weight for any of the stimuli (Table 4.22).

4.5.6 Overall effects of weight on responsiveness

No significant relationship between weight and novelty responsiveness was seen at any level of testing, refuting Hypothesis 4. This indicates that if there is a variable behavioural strategy at work underlying differences at an individual level, then it is not based immediately on body size. However, as for the other two variables being investigated, low calculated power indicates that the lack of significance cannot be definitively said to be because there is no relationship, due to the risk of Type II errors. As with the previous two variables, power analyses using a convention effect size rather than a calculated one give rather different results. In this case, power is generally raised to above 0.8, giving confidence that, in fact, the possibility of Type II errors has been avoided.

4.6 Sex, age and weight as possible confounding variables on individual variation in behaviour

In general and across all stimuli individually, there are no strong effects of sex, age or weight on behaviour. This thus means that there are few effects of an animal's age, sex or weight that would affect subsequent analysis of individual variation in responsiveness (Hypothesis 1, Section 1.4). Predictions made based on previous work are not supported, and in a minimum number of tests, males were more responsive to the food related stimuli. More differences and variations are seen significantly or near so for these stimuli than the others. This supports the conclusions in Chapter Three that, although kept in the analysis, the food related RIT and RIW should be assessed separately to the other novel objects.

Previous findings that females are more responsive to food or food-related stimuli (e.g. Box, 1988; Visalberghi *et al.*, 2003) were not supported in this study. It is possible that differences between this and previous studies, such as the method of stimulus presentation led to this difference. If faster response times of females are relative to males in the same group as them at the same stimulus presentation it will not be echoed in an individual presentation methodology as used here. If males are delaying approach

in deference to female group members, especially breeding partners, such differences in response will not be expected to be echoed during individual trials. The age and circumstance of the female marmosets being tested may also affect whether or not sex differences are seen. Breeding female marmosets can have two litters a year, so as adults will be almost constantly lactating or gestating. This will cause a high metabolic demand that requires a high level of food intake. Young females in single sex housing will not have these metabolic demands, and so a sex difference will not have been seen in the current study.

Throughout the analyses, power has been calculated by two methods. For all three of the variables investigated, and for all five of the behaviours measured, the low calculated power of the tests can be seen to be problematic. Where power is above 0.8, the recommended level for it to be set at, effect size is in the main less than 0.0001. This is effectively zero, and indicates that even with a sample size of 68, higher than many previous primate behavioural studies, it can be difficult to guarantee against the possibility of a Type II error. In the majority of tests, power was between 0.5 and 0.6, which is not much above a chance level for a Type II error to occur. Ironically, from the point of view of the larger analysis, it is the relatively large standard deviations, demonstrating a great amount of individual variation in response in the animals tested, that give a low effect size and thus a low calculated power. The very variability that is the subject of the larger analysis then assures a low power for any tests showing no significant differences or variation between or in traits.

The second method of power calculation relies on published conventions (quoted in Gpower; Buchner *et* al., 1993; Erdfelder *et al.*, 1996). If, rather than basing effect size on standard deviations or r^2 values, convention effect sizes are used (0.5 for sex differences, 0.15 for possible age weight relationships) satisfactory power is reached for almost all tests. The tests that do not reach a power of 0.8 are generally the two food related stimuli, that have in tests rather low alpha values. These low values make it more difficult to guarantee that there is no effect occurring in the data sample.

It seems reasonable on the basis of these results to assume that there are no effects of the three variables tested. It is important to be aware that due to the low power of some of the tests, there may be small effects that have been overlooked. Further investigations in the areas of sex and age differences in relation to novelty responsiveness should not

(and will not) be discarded on the weight of the current findings. If sex and age are included as potential covariates in the heritability analyses of behavioural traits, then any potential affect they might have can be taken into account. This is the approach taken in Chapter Six. The next Chapter, Chapter Five, investigates distilling behavioural response scores into a general responsiveness score, to be used in these heritability analyses.

Chapter Five

Developing measures of responsiveness

5.1 Introduction

Variability in animal responsiveness can be measured at several levels, from chemical, to physiological, to behavioural. At each of these levels of assessment, different measures may best represent the responsiveness of an animal. Different specific response variables may be appropriate in measuring different stimuli that an animal is responding to. For instance, heartbeat or heartbeat variation may best represent a physiological response to a novel object or situation in horses (*Equus caballus*) (Visser *et al*, 2002). Either changing colour or moving away may be the most representative response to threat in octopuses (*Octopus rubescens*) (Mather & Anderson, 1993).

Many studies of response in primates measure physical behavioural reactions (Table 5.1). Latency and duration scores are the most popular measures of the selection of studies listed in Table 5.1. In 16 of these studies, more than one measure is recorded. Latency times include latencies to approach (proximity), touch (or contact) and manipulate a stimulus, as well as feeding when appropriate. Duration scores include close proximity to, contact with and exploration of a stimulus. Studies also take note of spatial position, attention or head orientation when there is a lack of physical interaction. Sniffing, licking and contact with the nose are noted in five investigations. There are then many ways that response to a stimulus can be measured. In Chapter Three, the possibility that some behaviours may be more appropriate or informative was discussed. Is it also possible that the different variables recorded are reflective of an underlying behaviour or trait?

If there is a single underlying behavioural responsiveness that these different measures are tapping into, then data from those measures could be used to describe it. Scores from different behavioural variables could be combined or reduced down to one or two representative measures that adequately describe the variability seen. Data could in

Table 5.1: A selection of studies broadly representative of the methods used in measuring response behaviours in primates. N= novel, O= object, E= environment F= food, C= conspecifics, MP= model predator (N/O/F indicates a novel foraging task).

Reference	Species	Stimulus	Response Measure
Fragazy & Mason (1978)	*Callicebus moloch*	N/O	Proximity, contact, three levels of investigation: 1. Looking 2. Non-specific body contact 3. Grasping with hand; "general" responsive behaviour.
Box (1988)	*Callithrix jacchus*	N/E & N/O	All behavioural occurrences (exact behaviours measured not defined))
Forster (1995)	*Callithrix jacchus*	N/E & N/O	Staring at objects, handling objects, touching objects with nose, vocalisation, locomotion, chewing perch
Vitale *et al.* (1997)	*Callithrix jacchus*	N/O	Contact (mouthing, touching, sitting inside, sitting on top)
Rogers (1999)	*Callithrix jacchus*	N/E	Latency to leave nest box, number of leaps, movement of head, head cocking, touches of objects and numbers of objects touched
Rogers (1999)	*Callithrix jacchus*	N/O	Number of periods on object platform
Blackwood (2000)	*Callithrix jacchus*	N/O/F, N/O	Latency to approach, sniff and manipulate
Barros *et al.* (2000)	*Callithrix penicillata*	MP	Exploratory behaviour (smell, lick, scent mark, scratching, locomotor behaviour)
Day *et al.* (2003)	*Callithrix spp.*	N/F, N/O/F	Proximity (<50cm), latency to contact, manipulation and feeding; attentiveness
Majolo (2003)	*Callithrix jacchus*	N/O, N/O/F	Latency to explore, duration of exploration, latency to feeding, aggressive behaviours
Menzel & Menzel (1979)	*Saguinus fuscicollis*	N/O	Approach (<50cm), physical contacts
Hardie & Buchanan-Smith (2000)	*Saguinus fuscicollis*	N/O	Latency (enter area, first approach (<15cm), first touch)
Box (1988)	*Saguinus labiatus*	N/E & N/O	Scent marking, allogrooming, autogrooming, huddling, proximity, locomotion, manipulation, inspection, approach
Hardie & Buchanan-Smith (2000)	*Saguinus labiatus*	N/O	Latency (enter area, first approach (<15cm), first touch)
Box (1988)	*Saguinus oedipus*	N/E & N/O	Scent marking, allogrooming, autogrooming, huddling, proximity, locomotion, manipulation, inspection, approach
Day *et al.* (2003)	*Saguinus spp.*	N/F, N/O/F	Proximity (<50cm), latency to contact, manipulation and feeding; attentiveness
Day *et al.* (2003)	*Leontopithecus spp.*	N/F, N/O/F	Proximity (<50cm), latency to contact, manipulation and feeding; attentiveness
Fragazy *et al.* (1997)	*Cebus apella*	N/F & N/O	Interest (interaction with conspecifics regarding food), picks up food, eats food
Visalberghi *et al.* (2003)	*Cebus apella*	N/F & N/O	Visual exploration (<1m), contact (touching and holding), manipulation (active interaction), sniff-licking (holding), eating
Fragazy & Mason (1978)	*Saimiri sciureus*	N/O	Proximity, contact, three levels of investigation (1. Looking 2. Non-specific body contact 3. Grasping with hand), "general" responsive behaviour.

Table 5.1 Continued

Fairbanks (2001)	*Cercopithecus aethiops sabaeus*	C	Latencies (approach, touch, sniff, threaten), spatial position and head orientation, anxiety related and non directed expressions of arousal
Alexander & Hines (2002)	*Cercopithecus aethiops sabaeus*	O (toys)	Durations (approach to <2m, contact)
Capitanio (1999)	*Macaca mulatta*	Video of C	Durations (looking, withdrawal), aggressive and threatening behaviours
Joubert & Vauclair (1986)	*Papio hamadryas papio*	N/O	Looking, sniffing, touch with hand, grasp, transport (hierarchical complexity)
Hopkins & Bennett (1994)	*Pan troglodytes*	O	Latency to approach

effect be distilled into a purer description of a response continuum. Methods of reducing the dimensionality of data include Principal Component analysis (PCA) and Factor Analysis (FA). This chapter looks at the possibility of producing a "responsiveness" measure from the various behavioural variables recorded during observations in this study.

5.1.1 Analysis using Principal Component Analysis and Factor Analysis

Principal component analysis (PCA) and Factor analysis (FA) are two alternative but closely related techniques that transform data into a set of components or factors based on underlying variation. Where it is possible that several measures may be describing very similar, if not the same, underlying variables, they can be used to (attempt to) reduce the number of variables in an investigation. Quinn & Keough (2002) suggest that in biological research PCA is the more used of the two, because it is the more appropriate with regard to the subsequent use of the newly created variables. They summarise the difference between the two methods (based on Jackson, 1991) as PCA "trying to extract components that explain the variability in the original variables" and FA "trying to explain correlations among the original variables" (p459). What does this mean biologically?

Both methods of analysis have been used in assessment of the behaviour of primates and other animals. Stevenson-Hinde and colleagues pioneered the technique of using PCA to assess individual behavioural variation in primates in their studies of rhesus macaques (Stevenson Hinde & Zunz, 1978, Stevenson-Hinde *et al.*, 1980). In these studies, animals were observed and assessed using behaviourally defined adjectives.

Using a seven-point ratings scale they identified twenty-three personality and temperament traits that were reliably scored across several observers. Using principal component analysis (similar to factor analysis as used by Chamove *et al.* (1972) in a study of rhesus macaques) three separate, independent, dimensions ("confident-fearful"; "active-slow" and "sociable-solitary") were derived from these traits (Stevenson-Hinde & Zunz, 1978). In a subsequent paper, Stevenson-Hinde *et al.*, (1980) describe dimensions derived over three years of observation as "confident", "excitable" and "sociable". The first component in each of these studies ("confident-fearful" or "confidence") is based around traits of dominance and aggression (Bolig *et al.*, 1992). Ratings of these dimensions are reported as reliable and consistent across observers, and also predictive of some non-social behaviour (Clarke & Boinski, 1995). Using the technique to investigate the individual distinctiveness of cats, *Felis sylvestris cattus,* Feaver *et al* (1986) found that "[t]he results of the ratings and direct methods (of behavioural measurement) were significantly correlated" in five out of six cases. Using the same method as Stevenson-Hinde and Zunz (1978), but replacing principal component analysis with factor analysis, Capitanio (1999) reports some consistency across time in the personality dimensions of his macaque subjects. He also adds a fourth personality dimension, "equability", to those produced by Stevenson-Hinde and co-workers. Bolig *et al.* (1992) use this same method in combination with objective behaviour measurements to investigate personality and reactivity in rhesus macaques. In this study, as in Capitanio's (1999), principal component analysis of the personality traits measured reveals four major components rather than three, and the first principal component traits are related to reactivity and response rather than aggression and dominance traits. Reactivity was found to be correlated (either positively or negatively) with 10 of 20 reliably measured personality traits. "Apprehensive", "excitable", "fearful", "insecure", "irritable" and "tense" were positively correlated with reactivity, whereas "confident", "curious", "equable" and "understanding" showed negative association. The authors suggest that as little as three personality trait measurements may be enough to assign a reactivity level to all subjects. Consistency in the measurement of reactivity levels, however, is not total, as there is only complete agreement from raters of reactivity (on a scale of one to three) with 7 of the 22 animals in the study (32%).

There are some indications then, that even though this approach is useful and valid in investigating personality and how to rate it in primates, it is not always consistently and

completely reliable. The results of Capitanio (1999) and Bolig *et al.*'s (1992) studies do differ slightly from each other and those of Stevenson-Hinde (Stevenson-Hinde *et al.*, 1978, 1980). There is within these results either some lack of consistency between the workers assessing the animals, or a difference between groups of conspecifics in what can reliably be called a personality trait. Some of these potential problems are avoided in the current study. Rather than assessing behavioural traits based on adjectives derived through observation, the current study is based on specific quantifiable behaviours. This removes subjectivity from observer assessment. As only one group of animals was studied, it is not possible to investigate cross population variation in the responsiveness measured. A recent study of rhesus macaque behaviour (Williamson *et al.*, 2003) used FA to group behaviours related to fear and anxiety, in a way similar to the current study. Measures across testing paradigms produced seven factors: distress vocalisations, movement, distress cues, delayed independence, early independence, explore familiar environment and explore novelty. As indicated by their names, these factors grouped behavioural measures by test and by type of behaviour.

In Chapter Three, novel stimuli were categorised both by their own appearance and the general reaction of animals to them. These categorisations can be used when trying to find a general measurement for response in the animals tested. Stimuli within one categorization may be expected to elicit response along one continuum, and so the measurements of responsiveness could be reduced to reflect this.

5.2 Methods

In practice, PCA solutions to problems of data reduction often are very similar to FA solutions to the same problem. "The choice of common factors or components methods often makes virtually no difference to the conclusions of a study" (Cliff, 1987). PCA analyses variation across an entire data set. FA splits the contribution made by any variable in an analysis into a common and a unique component, and only uses the common component in the analysis. The sum of the common components is called the communality. Following the majority of primate studies, and acknowledging that all variation of the data is potentially important in the study, PCA was used to investigate any underlying relationships in response measures.

In a Principal Component Analysis, the components produced are the best unrelated explanations of the variance in the overall data set. Component I explains the most variation in the data and is a line of best fit along the axis of most variation. Component II is completely uncorrelated with the first, explaining the variance at right angles in the data spread. Subsequent components explain remaining, smaller amounts of variance. Eigenvalues are used to represent the amount of the original variance that each of the new derived variables (i.e. the components) explains. Components with Eigenvalues less than 1 are of less importance statistically as they only describe a very small proportion of the variance. This is used as an (arbitrary) cut-off point to avoid dealing with a lot of components that have little effect. Loadings for each of the new components for each original variable give the amount of the variance of that variable that the component represents.

All analyses were carried out using SPSS version 11.0.1. Because such data reduction in SPSS runs through a Factor Analysis program, it is essential to set the analysis to produce as many components or factors as input variables. This is so that the communalities of all variables equal one, and all variation across the original measures will be involved in the analysis.

Two series of analyses were carried out on the data. The first series, henceforth referred to as PCA-IND, analyses data from individual stimuli in the sets given by the categorisation in Chapter Three. Analyses for the five behavioural measures were thus carried out for individual stimulus scores in the following four groups: mirror (MIR), unattractive stimuli (MOU, EGG), food-related stimuli (RIT, RIW), and novel stimuli (TUB, POT, SYR, BAL). The second series, henceforth referred to as PCA-MEAN, analyses mean scores from the stimuli in each stimulus group. Four analyses were carried out using means scores for the stimuli within the above categories.

5.3 PCA-IND: results for individual stimulus scores grouped by stimulus category

The first series of PCA carried out looked at reactions to the stimuli split into the categories given in Chapter Three. Each analysis examines all five behavioural measures for each stimulus in the category.

Table 5.2a: PCA-IND Components derived from a PCA of response variables for MIR. Data in bold refers to components with an Eigenvalue of one or over. Total shows the Eigenvalue for each component; % of variance the amount of variance across the whole sample explained by that component. Cumulative % shows the total % of the variance explained by adding each consecutive component together.

	Initial Eigenvalues		
Component	Total	% of variance	Cumulative %
I	**2.231**	**44.627**	**44.627**
II	**1.572**	**31.442**	**76.069**
III	0.603	12.067	88.135
IV	0.391	7.820	95.956
V	0.202	4.044	100.000

Table 5.2b: PCA-IND Component Matrix. Loadings for components derived from PCA of response variables for MIR. Each loading gives the amount of the variance of that original variable that the component represents (see 5.2). ATT refers to attendance; TC to total time spent in contact; TP to total time spent in proximity; AL to latency to approach and CL to latency to contact. Only components with Eigenvalues of one or over are shown

	Component	
Variable	I	II
ATT MIR	0.534	0.681
TC MIR	0.784	0.262
TP MIR	0.926	0.136
AL MIR	-0.463	0.708
CL MIR	-0.510	0.721

5.3.1 PCA-IND: mirror

The mirror (MIR) is a stimulus categorised separately to the other objects because of its unique reflective properties and the animals' reaction to it (Chapter Three). A principal components analysis was carried out on the five measurements of response to the mirror for all subjects. This was to ascertain whether a dimensionally reduced score would display an underlying essence of response. The first two components in the analysis have Eigenvalues of above 1, and together explain 76% of the variance seen in the scores (Table 5.2a). Component I has relatively high loadings for total time spent in proximity and contact (Table 5.2b). As might be expected, both latency times load negatively. This first component thus describes animals' time spent in close interaction with the mirror, and explains 45% of the variance seen. Attendance and both latency scores load highly onto Component II. This component thus describes animals' latency

to interact with the mirror, as well as incorporating something of general attendance (which loads relatively high on both components).

Table 5.3a: PCA-IND Components derived from a PCA of response variables for MOU and EGG. See Table 5.2a.

	Initial Eigenvalues		
Component	Total	% of variance	Cumulative %
I	**4.538**	**45.381**	**45.381**
II	**2.195**	**21.954**	**67.336**
III	**1.064**	**10.643**	**77.978**
IV	0.707	7.071	85.050
V	0.542	5.424	90.474
VI	0.302	3.024	93.498
VII	0.258	2.579	96.077
VIII	0.152	1.523	97.600
IX	0.135	1.348	98.947
X	0.105	1.053	100.000

Table 5.3b: PCA-IND Component Matrix. Loadings for components derived from PCA of response variables for MOU and EGG. See Table 5.2b.

	Component		
Variable	I	II	III
ATT MOU	0.425	-0.087	0.717
ATT EGG	0.631	0.425	0.417
TC MOU	0.650	-0.469	0.129
TC EGG	0.782	0.408	-0.025
TP MOU	0.748	-0.567	0.083
TP EGG	0.827	0.454	0.035
AL MOU	-0.630	0.594	0.291
AL EGG	-0.579	-0.419	0.407
CL MOU	-0.705	-0.468	0.259
CL EGG	-0.673	0.588	0.181

5.3.2 PCA-IND: "unattractive" stimuli.

In Chapter Three, the mouse and cheese toy, MOU, and the cardboard egg tray, EGG, were categorised together as "unattractive" stimuli, although there were differences between them in how animals generally reacted. A principal components analysis of the five response measures for the two stimuli shows three components with Eigenvalues above one (Table5.3a). These components together explain approximately 80% of the variance seen. As with the mirror, the first component describes time spent near the stimuli (Table 5.3b). It has high loadings for total contact and total proximity and high negative loadings for latency times across both objects. The second component has moderate loadings, both positive and negative (0.4-0.594) for all variables except visual

attendance to MOU, which is very low (-0.0087). The loadings in each measure of response are in opposite directions for the two stimuli. This second component thus, if anything, contrasts MOU and EGG. The third component has a high loading for visual attendance to MOU (0.717), something lacking in the first two. The second highest loading is for visual attendance to EGG (0.417). The uniquely high loading for visual attendance to MOU may reflect its properties as a stimulus. The mouse head itself, being the same size as that of a marmoset, eliciting a definite response from the animals in the form of looking, but a wariness to approach or touch it.

5.3.3 PCA-IND: food related stimuli

A principal component analysis of the responses measured to the two food-related stimuli, RIT (raisin in a sealed tube) and RIW (raisin in a water bath), produces three components with Eigenvalues over one (Table 5.4a). Together these components explain 73% of the variance. As with the two previous analyses, the first component

Table 5.4a: PCA-IND Components derived from a PCA of response variables for RIT and RIW. See Table 5.2a.

	Initial Eigenvalues		
Component	Total	% of variance	Cumulative %
I	**3.305**	**33.052**	**33.052**
II	**2.369**	**23.694**	**56.746**
III	**1.629**	**16.289**	**73.035**
IV	0.828	8.283	81.318
V	0.579	5.790	87.108
VI	0.459	4.588	91.696
VII	0.339	3.389	95.085
VIII	0.268	2.676	97.761
IX	0.148	1.482	99.243
X	0.075	0.757	100.000

Table 5.4b: PCA-IND Component Matrix. Loadings for components derived from PCA of response variables for RIT and RIW. See Table 5.2b.

	Component		
	I	II	III
ATT RIT	0.737	-0.247	0.519
ATT RIW	0.677	0.573	-0.111
TC RIT	0.729	-0.440	0.283
TC RIW	0.585	0.649	-0.139
TP RIT	0.785	-0.369	0.332
TP RIW	0.716	0.585	-0.223
AL RIT	-0.236	0.697	0.331
AL RIW	-0.362	0.219	0.459
CL RIT	-0.217	0.588	0.448
CL RIW	-0.255	0.085	0.753

Table 5.5a: PCI-IND Components derived from a PCA of response variables for TUB, POT, SYR and BAL. See Table 5.2a. Only the first 10 of 25 components are shown.

	Initial Eigenvalues		
Component	Total	% of variance	Cumulative %
I	**7.084**	**35.419**	**35.419**
II	**3.394**	**16.972**	**52.391**
III	**1.891**	**9.456**	**61.847**
IV	**1.444**	**7.221**	**69.068**
V	**1.288**	**6.440**	**75.508**
VI	**1.045**	**5.226**	**80.734**
VII	0.910	4.548	85.282
VIII	0.572	2.858	88.140
IX	0.527	2.637	90.777
X	0.316	1.579	92.356

Table 5.5b: PCA-IND Component Matrix. Loadings for components derived from PCA of response variables for TUB, POT, SYR and BAL. See Table 5.2b. Only components with an Eigenvalue above 1 are shown.

	Component					
	I	II	III	IV	V	VI
ATT TUB	0.595	0.272	0.045	0.166	0.349	0.015
ATT POT	0.442	0.674	0.258	-0.059	-0.174	-0.012
ATT SYR	0.583	0.475	0.112	0.036	-0.133	0.511
ATT BAL	0.606	0.409	-0.333	0.213	0.120	-0.082
TC TUB	0.603	-0.260	0.315	0.487	0.291	-0.059
TC POT	0.672	0.388	0.276	-0.193	-0.117	-0.263
TC SYR	0.759	0.156	0.029	-0.207	-0.055	0.386
TC BAL	0.698	0.148	-0.394	0.317	0.154	-0.120
TP TUB	0.658	-0.393	0.356	0.407	0.218	-0.045
TP POT	0.723	0.408	0.322	-0.111	-0.146	-0.270
TP SYR	0.800	0.231	0.007	-0.193	-0.092	0.337
TP BAL	0.662	0.179	-0.507	0.233	0.017	-0.196
AL TUB	-0.358	0.645	-0.178	-0.149	0.139	-0.265
AL POT	-0.661	0.400	-0.205	0.251	0.213	0.266
AL SYR	-0.550	0.337	0.282	0.503	-0.389	-0.095
AL BAL	-0.500	0.315	0.563	-0.003	0.362	0.081
CL TUB	-0.317	0.713	-0.274	-0.229	0.208	-0.203
CL POT	-0.549	0.384	-0.283	0.284	0.169	0.338
CL SYR	-0.475	0.466	0.155	0.443	-0.513	-0.030
CL BAL	-0.411	0.374	0.479	-0.104	0.468	-0.031

describes time spent near the stimuli, with high loadings for both stimuli across visual attendance, total proximity and total contact (Table 5.4b). All of these loadings are above 0.65 except that for total time spent in contact with RIW, which is 0.585. The second component contrasts between the stimuli the three response variables that load highly together in Component I. Total contact, total proximity and visual attendance show negative loadings for RIT and positive ones for RIW. High loadings are also given

to latency times for RIT, in the opposite direction to the duration times. The positive loadings across the stimuli indicate either a connection between duration scores for RIT and latency scores for RIW, or more likely, the contrast between latency and negative duration scores for RIW. Also, as with the analysis of unattractive stimuli, contrasts between the stimuli are being demonstrated. Component three shows the highest loading for latency to come into contact with RIW (0.753), and moderate loading for the other latency scores. Total visual attendance for RIT has the second highest loading, 0.519. If anything then, component three describes to some extent the latency scores seen for both stimuli.

5.3.4 PCA-IND: "novel" stimuli

The final category described after analysis in Chapter Three is that of novel stimuli. These are objects without significant unique features, such as food or reflections, and with a similar range of responses from the marmosets. The category includes the grey corrugated plastic tube (TUB), the blue pot (POT), the syringe (SYR) and the hollow pink ball (BAL). A principal components analysis of the five response measures for these four stimuli has six components with Eigenvalues of over one (5.5a). Together, these six components explain over 80% of the variation in the data. Component I, which explains 35% of the variation seen shows moderate to high loadings for visual attendance and total proximity and contact scores (0.442-0.800), along with negative latency loadings from –0.317 to –0.661 (Table5.5b). As with previous analyses this component seems to be explaining an axis of visual and physical attention to the stimuli that contrasts with latency times to both approach and touch. The other five components with Eigenvalues of over one do not seem to demonstrate obvious patterns. For Component II, the highest loadings are for visual attendance to POT, latency to approach and contact TUB. Component three has only two variables that have a loading of over 0.5, which are total time spent in proximity to BAL (-0.507) and latency to approach BAL (0.563). The next highest loading, 0.479, is latency to approach BAL. With all other loadings below 0.4, this component then seems to describe reaction to the stimulus BAL. For Components 4, 5 and 6, there is no discernable pattern in the higher loadings, be they positive or negative. The components explain 7.2%, 6.4% and 5.2% of the total variance respectively.

5.3.5 Interpreting PCA-IND

Throughout the PCA analyses a contrast between total contact and proximity times and latency times can be seen. In the analysis of response to MIR, total time scores load highly on Component I, and latency times load highly on Component II. This gives two components that usefully describe response to the stimulus with reduced dimensionality compared to the original measures. The first component of the PCA for MOU and EGG contrasts the different measures, with high positive loadings for total times and high negative loadings for latencies. Analysis of food containing stimuli (RIT and RIW) shows a first component explaining variation in total time measures. Latency times, however, are not thus characterised. High positive loadings for latency times are seen in Component II, but they are shared with several total time measures. The fourth set of objects, the novel stimuli, do show a contrast between latency times and total time spent in proximity and contact, notably in Component I. Here, although not all loadings are high, all durations are positive, and all latency times negative.

These analyses thus show common underlying threads to measures for most of the stimuli. The contrast between latency times and duration times is an obvious one, and does not need a principal component analysis to elucidate it. What a principal component analysis can do is help to quantify it. Where a single component highlights and contrasts the measures; and where a large proportion of the variance is explained by that component, it could be used to show a general response to the stimuli tested. Within these analyses however, the first component never explains more than 45% of the variance. A higher percentage of the variability in the data needs to be explained to make any PCA useful. The analysis of the novel stimuli needs so many components to explain the variability seen that it has little advantage over the original scores themselves.

5.4 PCA-MEAN: results for mean stimulus reaction scores for each stimulus category

The second series of PCA examined the five behavioural scores for each stimulus category (as opposed to individual stimuli within each category, as above). The individual animals' scores used were means for each response behaviour calculated across the stimuli in the category.

5.4.1 PCA-MEAN: mirror

As the mirror was placed in its own category in Chapter Three, the PCA is identical to that in 5.3.1. In this analysis, two components describe variance in duration times and latency times respectively. Together the components explain 76% of the variance.

5.4.2 PCA-MEAN: unattractive stimuli

A PCA of mean response scores for the two unattractive stimuli, MOU and EGG, shows only one component with an Eigenvalue of one (Table 5.6a). This component loads highly for all five measures, the lowest being 0.630 for time spent in visual attendance, to 0.946 for time spent in proximity (Table 5.6b). Both latency scores have high negative loadings, which contrasts with the high positive loadings for duration times. This first component explains approximately 70% of the variance seen across the data.

A second component, explaining a further 16.8% of the variance has the highest loading for visual attendance at 0.710, and all other loadings are 0.485 (mean latency to approach) and below. Its Eigenvalue, however, is below one, which indicates it is not a statistically important component in the analysis. The first component then is an adequate descriptor for the variation seen in response to these unattractive stimuli,

Table 5.6a: PCA-MEAN Components derived from a PCA of five mean response variables to MOU and EGG. See Table 5.2a.

	Initial Eigenvalues		
Component	Total	% of variance	Cumulative %
I	**3.499**	**69.979**	**69.979**
II	0.840	16.808	86.787
III	0.373	7.461	94.249
IV	0.173	3.462	97.711
V	0.114	2.289	100.000

Table 5.6b: PCA-MEAN Component Matrix. Loadings for components derived from PCA of mean response variables to MOU and EGG. See Table 5.2b.

	Component
Variable	I
ATT	0.630
TC	0.885
TP	0.946
AL	-0.788
CL	-0.896

Table 5.7a: PCA-MEAN Components derived from a PCA of five mean response variables to RIT and RIW. See Table 5.2a.

	Initial Eigenvalues		
Component	Total	% of variance	Cumulative %
I	**2.656**	**53.124**	**53.124**
II	**1.359**	**27.180**	**80.304**
III	0.606	12.115	92.418
IV	0.239	4.777	97.195
	0.140	2.805	100.000

Table 5.7b: PCA-MEAN Component Matrix. Loadings for components derived from PCA of mean response variables to RIT and RIW. See Table 5.2b.

	Component	
Variable	I	II
ATT	0.872	0.328
TC	0.903	0.089
TP	0.942	0.126
AL	-0.343	0.779
CL	-0.277	0.788

contrasting duration times and latency times to explain most of the variance seen in the data.

5.4.3 PCA-MEAN: food related stimuli.

A PCA of the means of the five response measures used for the two food-related stimuli (RIT and RIW) shows two components with Eigenvalues of one (Table 5.7a). Together these two components explain over 80% of the variation seen in the data. The first component explains 53% of the variance in the data and has loadings of over 0.9 for both duration scores (Table 5.7b). Visual attendance has a loading of 0.872. Both latency scores have low, but negative scores, contrasting them slightly with the duration scores. The second component, explaining a further 27% of the variance shows relatively high loadings for latency scores (0.779 for mean latency to approach, and 0.788 for mean latency to contact), and low scores, 0.328 and below, for the three other measures. The two components thus together give good representation of the general response to food related stimuli. Component I describes variation in how long animals stay close to and in physical contact with stimuli; Component II describes variation in how long it takes animals to initially approach and touch the stimuli.

5.4.4 PCA-MEAN: novel stimuli

Principal component analysis of mean response scores to the four "novel" stimuli (TUB, POT, SYR and BAL) gives two components with Eigenvalues of over one (Table 5.8a) Component I explains 63% of the variance in the data and has an Eigenvalue of 3.172. The mean duration times load most highly, with mean contact duration at 0.893 and mean proximity duration at 0.939 (Table 5.8b). Attendance is also relatively high, at 0.661. Both mean latency scores have reasonably high negative loadings (approach -0.760, contact –0.692). A second component, with an Eigenvalue of 1.335, explains 26.7% of the remaining variance in the data. This component has relatively high loadings for attendance (0.675) and both latency to approach (0.569) and latency to contact (0.661). The duration loadings are much lower (proximity 0.242, contact 0.243), although unlike in Component I, in the same direction of the latency scores. As with the unattractive stimuli then, Component I is an adequate descriptor for the variation seen in response, contrasting duration times and latency times to explain most of the variance seen in the data. Component II, however, also explains a large proportion of the variance, and indicates a link between attention and latency times.

Table 5.8a: PCA-MEAN Components derived from a PCA of five mean response variables to TUB, POT, SYR and BAL. See Table 5.2a.

	Initial Eigenvalues		
Component	Total	% of variance	Cumulative %
I	**3.172**	**63.433**	**63.433**
II	**1.335**	**26.698**	**90.131**
III	0.245	4.899	95.030
IV	0.160	3.208	98.238
V	0.088	1.762	100.000

Table 5.8b: PCA-MEANComponent Matrix. Loadings for components derived from PCA of mean response variables to TUB, POT, SYR and BAL. See Table 5.2b.

	Component	
Variable	I	II
ATT	0.661	0.675
TC	0.893	0.243
TP	0.939	0.242
AL	-0.760	0.569
CL	-0.692	0.661

5.4.5 Interpreting PCA-MEAN

To summarise, the second series of analyses produced meaningful components that explain a much larger proportion of the variance in the sample than PCA-IND. Results for the mirror are identical to those for PCA-IND, with response characterised by two factors representing duration of interest and latency to close interaction. Response to the two "unattractive" stimuli can be characterised by a single factor that shows latency to interaction contrasted against duration of interaction and explains approximately 70% of the variance seen in the data. The animals' reaction to the two stimuli containing food can be assessed using two scores derived from the PCA. As for MIR, the first describes the amount of time spent near or touching the stimulus, the second describes latency to interaction. Variation in response to the four "novel" stimuli can be described by two scores. The first describes the contrasting duration and latency scores. The second score is not as straight forward as the other components in this series of analyses, describing as it does variation in attendance and latency. Between the two scores they describe over 90% of the variation in response to the four novel stimuli. Table 5.9 summarises the useful components and shows how much of the variance they describe.

Throughout the analyses, visual attendance stands out as a measure not following the consistent pattern that latency scores and duration times seem to. Instead, it tends to load relatively highly (0.5 or above) with ALL components having an Eigenvalue of more than one. The exception to this is a loading of 0.328 for Component II of the food related PCA, which describes latency scores rather well. This could be because it is the only measure that does not require physical closeness to the stimulus. Therefore, if a

Table 5.9: Summary of PCA-Mean results suitable for describing individuals' reaction to the stimulus groups

Stimulus Group	Component (with Eigenvalue ≥1)	Component describes:	% of variance described
Mirror	I	Duration	44
	II	Latency	31
Unattractive	I	Duration & Latency	69
Food related	I	Duration	53
	II	Latency	27
Novel	I	Duration & Latency	63
	II	Latency	27

stimulus is unattractive but noticeable it will elicit as high visual attendance score as will an attractive stimulus. With the other measures, unattractive stimuli will elicit a high latency score (or indeed a censored value) and low duration scores, and vice versa for an attractive stimulus. This means that visual attendance as a measure of reaction can cut across the contrasting latency and duration scores. From the perspective of individual animals' responses, whereas a responsive animal will have low latency and high duration scores, and a negatively responsive animal will have high latency and low duration scores, both can have high visual attendance scores. This gives visual attendance the potential to convey important information about animal response not recorded by proximity measures. What visual attendance does not do, however, is differentiate between attraction and repulsion (or caution) with regard to the stimulus. This means that in some ways, it conveys less information than the other measures. Unfortunately in this set of PCAs, it did not have enough effect on the general variance in response (or, in fact, had too consistent a loading on any component) to be helpful in analysis and interpretation.

5.5 Describing the response measures for use in a heritability analysis

The large amount of variance explained by the components of PCA-MEAN described in section 5.4.5 and in table 5.9 means that they can be used to represent responses to novelty. One of the major aims of this thesis is to investigate the possibility of a measurable genetic influence on variation in responsiveness in the laboratory population

---

Table 5.10: Selective statistics for six continuums derived from the PCA-MEAN analyses. R refers to responsiveness. Min.=minimum, Max.= maximum. K-S refers to a Kolmogorov-Smirnov test for normal distribution. Continuums with a p-values of ≤ 0.05 are not normally distributed. * refers to a minimum p-value. Other descriptive statistics can be found in the text.

Continuum	Range			Normal distribution (K-S)	
	Min.	Max.	Range	Z	p
Mirror duration continuum	-2.84	2.20	5.04	0.086	0.200*
Mirror latency continuum	-2.46	1.81	4.27	0.074	0.200*
Unattractive R continuum	-1.48	2.64	4.12	0.145	0.001
Food duration continuum	-2.90	2.32	5.22	0.120	0.016
Food latency continuum	-2.21	2.64	4.85	0.138	0.003
Novel R continuum	-2.82	2.10	4.92	0.67	0.200*

of marmosets. To do this a heritability analysis can be carried out on a trait, and this will be the focus of the next chapter. For the heritability analysis, the components produced by the PCA can be used for this analysis. If possible, it would be preferable for only one measure for each stimulus group to be used to investigate heritability in response, as it would mean that the responsiveness continuum was being represented by only one variable. MIR and the food related stimuli need two components each, one describing duration and one describing latency. This means that these two aspects of response need to be tested separately in the heritability analysis. In the PCA for the unattractive stimuli, a single component can be used to describe response incorporating both how long it takes an animal to approach and touch an object, and for how long it will continue interacting with it. The novel stimuli PCA, described two components that explained a large proportion of the variance seen. Component I, describes more than twice the amount of the variance than Component Two. Also, the first component describes and contrasts both latency and duration scores. Component II does load highly for latency scores, but not as highly as Component I. This questions whether or not Component Two of the analysis is useful, even though it does describe almost 27% of the variance seen in the sample. This is the same amount as component II of the PCA for the food related items, but that describes latency scores as opposed to the duration scores described by Component I in that analysis. Component I can be used alone, explaining the majority of the variance in scores and also describing and contrasting both latency and duration. For two sets of stimuli then, "unattractive" stimuli and "novel" stimuli, one component will be used to represent responsiveness in the heritability analyses of the next chapter. The mirror, MIR, and the food-related stimuli, both need two components to represent response. Interestingly, they also have more aspects to the stimuli other than "novelty" to respond to, namely, the reflective quality of the mirror, and the raisin reward in the tube and the water bath. Where a single component is used, it will be referred to as the responsiveness continuum for that stimulus group. Where two components are used, they will be referred to as the duration continuum and the latency continuum, as this best characterises the way they describe the data. All the score continuums to be used have a mean of 0, a variance and standard deviation of 1 and a standard error of 0.121. Table 5.10 describes the range of each continuum and whether it is normally distributed.

5.6 General conclusions

Both PCA-IND and PCA-MEAN provide component scores that can describe the animals' reaction to the stimuli. For PCA-MEAN, unlike PCA-IND, these components describe a high proportion of the variance in the data. This makes individual animals' scores for these meaningful components suitable for characterising their response to novelty of different forms. The original data from behavioural observations are thus reduced in dimensionality enough to be usefully used as behavioural measures in a behavioural genetic analysis to establish whether there is a genetic basis to variation in responsiveness in this population of common marmosets. The study differs from most previous uses of PCA in behavioural work that has sought to explain personality differences as here several measured behaviours have been reduced to one or two underlying components. In previous studies using similar techniques, results from the reduction of data from subjective behavioural descriptives were used to find underlying personality traits (Stevenson-Hinde & Zunz, 1978; Stevenson-Hinde et al., 1980; Bolig et al. 1992; Capitanio, 1999). Some recent studies, especially Williamson *et al.*, (2003), have used PCA in a similar way, and also carried out heritability analyses on the results as is described in Chapter Six of this study.

Chapter Six

The Heritability of Behavioural Traits Related to Response

6.1 Introduction

Previous chapters have described the behavioural variation in response to a set of stimuli in a group of common marmosets and response continua derived from the measurements used. Five behaviours related to response, visual attendance, latency to first approach, latency to first contact, duration of proximity and duration of contact were demonstrated to be measurable and variable across the animals tested. The nine stimuli used in the presentations were grouped based on these responses for further testing (Chapter Three). The age and weight of the animals were found not to have any major effect on the variation in response, but some small yet significant effects of sex were noted (Chapter Four). Principal component analyses were carried out on the stimulus groups defined. Components produced by the analysis of mean response scores to the stimulus groups were used to produce a set of general response continua (Chapter Five). This chapter investigates using heritability analyses to assess the genetic influence on these response continua produced and the mean stimulus group responses themselves.

6.1.1. Heritability

Heritability is essentially a statistic that estimates the genetic effect size on a trait (i.e. how much of the variation seen in the trait has a genetic basis), be it a physical characteristic (Roberts *et al.*, 1978), a physiological measure (Rogers *et al.*, 2004), or a psychological trait (Bouchard, 1994). As such, it explains the genetic contribution to variation in a trait across a population, rather than the phenotype of a single individual (Plomin *et al.*, 2000). Two different definitions of heritability are discussed in the literature, broad-sense heritability, and narrow sense heritability (e.g. Futuyma, 1998; Falconer & Mackay, 1996; Plomin *et al.*, 2000. These two definitions vary in both their use and usefulness. Broad sense heritability, also referred to as the "degree of genetic determination" (Falconer & Mackay, 1996, p123) relates to all sources of genetic

variance, whether the genes operate in an additive manner or not (Plomin *et al.*, 2000). Narrow sense heritability, which most literature refers to as simply heritability, "expresses the extent to which phenotypes are determined by the genes transmitted from the parents." (Falconer &Mackay, 1996, p123). It refers only to the proportion of the phenotypic variance that is explained by additive genetic effects. It is possible that genetic and environmental differences in behaviour can be confounded, for instance when two family groups occupy different "microenvironments" (Futuyma, 1998). This is difficult to control for in humans and in non-human primates when breeding and the location of animals cannot be controlled.

Narrow sense heritability is the meaning that is most important in this study. It is more relevant than broad sense heritability for practical purposes (including animal breeding), as it gives an indication of the extent to which a trait will "breed true". Henceforth, heritability will refer to narrow-sense heritability as described above.

6.1.2 Using heritabilities to study trait variation

Heritabilities are specific to the population in which they are measured, at the time of measuring and are not indicative of a universal, species wide truth. This is because both the effect of an environment and the genetic effect on a trait in a population could vary greatly in relatively few generations, or indeed within one generation. For instance, if a local environment stabilises, or homogenises, genetic variability will become more important in describing variability within a population. The less variation there is in an environment, the less effect environmental variation will have on a trait. It can be expected that the heritability of traits will increase as their effect on reproductive fitness lessens. This is because the larger the effect a trait can potentially have on an animal's fitness, the less opportunity there is for genetic variation in that trait; i.e. if a trait is subject to strong selection, the genes responsible will go to fixation, at which point all the variation will be due to environmental factors. This has been demonstrated in studies of the heritability of several characters in fruit flies (*Drosophila melanogastor)* (Roff & Mousseau, 1987), as well as in a range of wild species (Mousseau & Roff, 1987).

Heritability has been used extensively as a tool in human psychology, often in twin studies looking at behavioural or personality traits. Bouchard and Loehlin (2001)

provide an extensive review. They look at heritability in both a broad and a narrow sense, using data from studies based on several alternative, but often related, frameworks for the study of human personality. These heritability scores range from 0.11 for competence (on the NEO personality inventory) to more than 0.5 for the "Big Five Factors" in some studies. There is great variety seen then in the heritability of human behavioural factors as related to personality. The message seems to be from these data that the genetic effect on variation in human personality traits is more than nothing, but less than all. Measures of the heritability of "Extraversion" (from the Big Five), a personality trait that must at least be in part related to responsiveness, range from 0.49 to 0.57 (see Table III, Bouchard and Loehlin; 2001). Heritabilities of NEO Personality Inventory scores that might relate to responsiveness, facets of "Extraversion" and "Openness to Experience", include 0.36 for excitement seeking and 0.34 for openness to new actions (Jang *et al.*, 1998). It is possible that the underlying (biochemical or neurological) basis for such human personality traits and some non-human primate behavioural traits are homologous.

Heritability calculations have been used for many purposes on the traits of non-human animals. Classic investigations include those for useful characteristics in livestock, such as fat thickness, weight gain and litter size in pigs (*Sus scrofa*) (Smith, King & Gilbert, 1962; Strang & Smith, 1979); body weight and egg weight in chickens (Emsley *et al.*, 1977). Studies of experimental animals also examine the heritability of traits, such as tail length and body weight in mice (Rutledge *et al.*, 1973), and body size, egg production and even abdominal bristle number in fruit flies (Robertson, 1957; Clayton & Robertson, 1957). Analysis of behavioural traits of working animals has also been carried out. A recent study of hunting behaviours in Swedish flatcoated retrievers (*Canis familiaris*) found heritabilities of 0.1-0.4, and used factor analysis to reveal broader personality traits related to these behaviours (Lindberg *et al.*, 2004).

There have been limited studies of the heritability of behaviours in non-human primates. Weiss et al (2000) investigated the heritability of personality factors in chimpanzees, using the human "big five" factors, extraversion, agreeableness, conscientiousness, emotional stability, openness to experience, and an additional factor, discovered to be important for chimpanzees, dominance (King and Figueredo, 1997). From the factors measured, only dominance showed significant narrow-sense heritability. It should perhaps be noted that the dominance factor is described as a broad, continuous

personality dimension rather than the situation specific suite of behaviours normally described in the primate literature. Primates are often used as models for research into human psychology, especially in personality linked areas such as anxiety and depression, and drug and alcohol dependency. Recently, several monoamine metabolites related to the neurotransmitters, serotonin, dopamine and noradrenaline, which are in turn related to personality and individual variation in psychological traits in humans, have been found to be heritable in baboons (Rogers, *et al.*, 2004). Behaviours reflecting increased stress responses and behavioural inhibition have found to be heritable in rhesus macaques (Williamson *et al.*, 2003). Some traits produced by clustering these behaviours using factor analysis were also shown to be heritable. Tests in the study included reactions to a remote control car, a human intruder, and novel and familiar fruits. Several behaviours across the tests showed significant heritability, including latency to leave mother during a free play session, latency to inspect novel fruit, and also movement and exploration of a standard cage in a novel room. Only one of the factors produced, reflecting movement in the tests, was significantly heritable at an alpha of 0.05. If it is accepted that retesting the same pedigree for the different behavioural measures counts as multiple comparisons because of the repeated use of the data, a correction for multiple comparisons should be used. It is possible that that would lead to the heritability being non-significant, but it cannot be ascertained from the information given in the paper. A Bonferroni correction, for instance, assuming a standard alpha of 0.05, requires a p-value of 0.0071 or below for the seven factors identified to be significantly heritable (see section 4.3).

Results of the previous chapters have demonstrated that marmosets show a large variation in their response to novel object presentations. The behavioural measurements recorded during these presentations can be considered individually or, by using statistical techniques, reduced down to a more general score representing variation along a responsiveness continuum. It is possible to carry out heritability analyses on these scores to investigate the genetic role in establishing the variation seen in animals' responses.

6.2. Heritability analyses

6.2.1 Pedigree construction

In order to carry out a heritability analysis it is necessary to understand the genetic relationships among the individuals studied. In many studies, participants are chosen because of genetic relationships, such as twins, and so analyses can be quite straightforward. In the current investigation however, there was an opportunistic approach to testing individuals, and so specific genetic relationships between individuals could not be guaranteed. Also, although they twin, marmosets tend to have dizygotic offspring (Sussman, 2002). This means that a twin study typical of many human psychological approaches could not be used. Modern computational techniques allow individuals with a variety of genetic relationships to be used in an analysis, as long as the natures of the relationships are known so that relationship coefficients can be calculated.

Information on the individuals tested was collected from paper records from the breeding colony at Porton Down where all individuals were bred and housed. To establish a pedigree of relative depth, details of all breeding individuals in the colony over a 25-year period for which records existed were taken, as well as details for the individuals tested. The data were then entered into the PedSys database system (Dyke, 1989), which was used to construct detailed pedigrees. From these data, individuals not related to any of the animals tested were pruned, until a single, extensive pedigree linking all individuals was produced. In order to run the analysis, it was necessary to add records for unknown parents of individuals with one known parent in the pedigree. SOLAR (see 6.2.2, below), the program used to calculate heritability, requires all individuals in a pedigree to have both or no parents. After addition of these unknown parents and additional animals from different experiments, the pedigree for analysis stood at 603 individuals.

6.2.2. Calculating heritability

The computer package Sequential Oligogenetic Linkage Analysis Routines (SOLAR version 1.6.6; Almasy and Blangero, 1998) was used to estimate the heritability (h^2) of response traits. SOLAR carries out variance component-based quantitative genetic

analyses (using its *polygenic* analysis command) that calculate both heritability and its significance. For all analyses, age at testing in days and sex were included as possible covariates. Although Chapter Four demonstrated that neither had a significant effect on responsiveness, there were indications of some effect before correcting for multiple comparisons. This was not the case for weight at testing, so it was left out of any analyses. Covariate screening was used to determine the statistical significance of each covariate and the interaction between the two, and remove it from the final analysis if it fell below a significance level of 0.05.

Two sets of analysis were carried out. The first analyses were on the six response continua made from factors in the PCAs of Chapter Five. The second set were on the five individual response measures, using mean scores from the stimulus groups defined in Chapter Three. The six continua defined were mirror duration continuum, describing variation in the time spent close to or touching the mirror; mirror latency continuum, describing variation in the latency to approach and touch the mirror; unattractive stimuli continuum, describing variation in response to the mouse head and egg box; food duration continuum, describing variation in time spent close to or touching the two food related stimuli; food latency continuum, describing the variation in latency to approach and touch the two food-related stimuli, and novel stimuli continuum, describing variation in response to the novel stimulus group. The stimulus groups defined in Chapter Three and used in the analyses were the mirror (MIR), unattractive stimuli (MOU & EGG), food-related stimuli (RIT & RIW) and novel stimuli (TUB, POT, SYR & BAL). As each of these groups could be investigated using five different measures, twenty heritability analyses were carried out based on this set of scores. Of the twenty, trait scores for five were not suitable for analysis in SOLAR, due to low standard deviations, which make calculations inaccurate. For each of these five traits, scores were multiplied by a treatment value suggested by the SOLAR program.

As the heritability analyses use the same pedigree data every time, it could be considered that multiple comparisons are being conducted. If this is the case, and repeatedly using a relationship matrix of animals is the same as repeatedly using a variable such as the animals' age or sex, then a correction for multiple testing should be used. If analyses are considered in two different sets, derived continua and stimulus group measures, then different Bonferroni corrections should be calculated for each (see Chapter Four, section 4.3). A Bonferroni correction for the derived continuum analyses

gives a significance level, or α, of 0.0084 ($\alpha'' = \alpha/k = 0.05/6$, where k=no. comparisons). For the set of stimulus group measures, a Bonferroni correction gives $\alpha = 0.0025$ for significance ($\alpha'' = \alpha/k = 0.05/20$), where 0.05 is the standard α.

6.3 Results

6.3.1 Derived continua

Analyses of the six derived continua showed no significant heritability (Table 6.1). In fact, none of the continuum scores had even registered a non-significant heritability from these estimations. It was demonstrated in Chapter Four that sex and age had a minimal effect on an animal's response. They were however, included in the analysis as covariates, to make sure that there was no low, yet significant, effect. As Table 6.1 shows, neither of the two covariates, nor the interaction between them, had a significant effect. This means that all of the possible covariates were then screened out of the final models used to estimate the heritability of the six continuum scores tested.

6.3.2 Response scores by stimulus group

The five original response scores were also tested, placed together in four groups as defined in Chapter Four (see Table 6.2). One behaviour, latency to approach unattractive stimuli, is highly heritable (h^2=0.81,p=0.0045) before correction for multiple comparisons. Both latency to come into contact with the mirror (h^2=0.29) and

Table 6.1: Heritability scores for the six continua derived from PCA in Chapter Five. Any significant scores are shown in bold. Non-significant covariates are excluded from the analysis

Continuum	Heritability	p-value	Covariate p-value		
			Sex	Age	Sex*age
Mirror duration	0.0000	0.5000	0.4713	0.2442	0.2038
Mirror latency	0.0000	0.5000	0.7938	0.5807	0.8632
Unattractive stimuli	0.0000	0.5000	0.8658	0.5681	0.4356
Food duration	0.0000	0.5000	0.4996	0.7522	0.7522
Food latency	0.0000	0.5000	0.4396	0.7160	0.8469
Novel stimuli	0.0000	0.5000	0.6534	0.7277	0.7864

the amount of time spent in close proximity to the mirror (h^2=0.4913) were approaching significance, with p-values of less than 0.1. A Bonferroni correction for multiple comparisons, however, gives α=0.0025 rather than 0.05, at which point none of the heritability scores are significant. It can be seen in Table 6.2 that of the five behavioural measures, only approach latency had (non-significant) heritabilities of more than zero for all four of the stimulus groups.

Within this set of analyses, the covariate sex had a significant effect on the variance found in three of the traits tested. Latency to close contact with and duration of proximity to food-related stimuli showed a significant proportion of variance due to sex differences at the α=0.05 level ($p<0.0000$ and $p=0.0421$) respectively. In fact, almost a quarter of the variance (0.2422) seen in latency to close contact, and a tenth of that seen for duration of proximity (0.1124) seems to be due to sex. Another latency time, mean

Table 6.2: Heritability scores for the five behavioural measures, partitioned in stimulus groups (see 6.2.2). Any significant scores are shown in bold. Non-significant covariates are excluded from the analysis. Bracketed numbers next to a trait indicated the number of times trait scores were multiplied to fulfil minimum standard deviation requirements for SOLAR. AT=attendance, TP= total time spent in proximity, TC= total contact, AL= approach latency, CL= contact latency.

Trait	Heritability	p-value	Covariate p-value		
			Sex	Age	Sex*age
Mirror AT (4.4)	0.0000	0.5000	0.8516	0.4946	0.3117
Unattractive AT (5.1)	0.0000	0.5000	0.4964	0.4957	0.4607
Food related AT (5.6)	0.0000	0.5000	0.1563	0.1984	0.4808
Novel AT (6.1)	0.1581	0.2434	0.5227	0.0313	0.2373
Mirror TP	0.4913	0.0744	0.1017	0.2045	0.1324
Unattractive TP	0.0000	0.5000	0.6487	1.0000	0.6985
Food related TP	0.0000	0.5000	**0.0421**	0.3355	0.4564
Novel TP	0.0000	0.5000	0.7846	0.3106	0.5224
Mirror TC	0.1824	0.2297	0.9259	0.2004	0.1863
Unattractive TC	0.0000	0.5000	0.3288	0.7611	0.7611
Food TC	0.3172	0.1223	0.1449	0.4710	0.8766
Novel TC	0.0000	0.5000	0.5339	0.4958	0.8441
Mirror AL	0.2086	0.1249	0.5105	0.6533	0.2353
Unattractive AL	**0.8089**	**0.0046**	0.9264	0.4051	0.1882
Food related AL (2.3)	0.1772	0.2290	0.5363	0.5363	0.1381
Novel AL	0.0394	0.4359	0.0831	0.1186	0.1847
Mirror CL	0.2906	0.0670	0.5266	0.7424	0.3380
Unattractive CL	0.2531	0.2759	0.4484	0.6735	0.9208
Food related CL	0.0000	0.5000	**0.0000**	0.2916	0.4661
Novel CL	0.0000	0.5000	**0.0091**	0.3237	0.5944

latency to close contact with the novel stimulus group was also significant at the α=0.05 level (p=0.0091). The proportion of the variance explained by sex in latency to close contact with novel stimuli was 0.0756, so almost 8% of the overall variation in scores. If, however, the same Bonferroni correction for multiple comparisons is applied as was for the significance of the final heritability scores in this set of analyses, sex only remains a significant factor in the variance of latency to close contact with food-related stimuli. As such, the influence of sex across the analyses can probably be assumed to be operating at a chance level

6.4 Heritability and behavioural traits

Across the two sets of heritability analyses no behavioural traits were found to be significantly heritable once controls for multiple comparisons were carried out. Only one behaviour, latency to approach unattractive stimuli, was found to be highly heritable without this Bonferroni correction (at α=0.05). If this was the only significant heritable behaviour, it would suggest a different mechanism of response to unattractive stimuli than to novel ones. This is implicit in the fact that the unattractive stimuli were considered so as they elicited a generally negative reaction, unlike the other stimuli. The case is strengthened by the significant heritability in reaction. It does not mean that reaction to unattractive stimuli is more important than the other responses measured as that might in fact lead to less genetically based variation. In order to untangle this, it would be necessary to understand what exactly about the stimuli the marmosets found unattractive, and whether it was a common factor, or different for each of the two stimuli in the group.

Of the five behavioural measures, only approach latency showed (even non-significant) heritabilities for all of the stimulus groups. All of the other response measures had at least two heritabilities of zero. This indicates that, of all the measures, latency to first approach to a novel stimulus is the most suitable measure to use when assessing the possibility of heritability in response to any stimulus.

6.4.1 Response to mirrors

The time it took animals to touch the mirror, and the amount of time they spent near it within a presentation session, did have p-values of below 0.1. This may be an indication

of, if similar tests were carried out with a larger sample size, a role for genetics in how animals react to mirror images (however the animals themselves actually perceive this image). It is possible that marmosets do not recognise their own reflections as themselves, so the reaction to a mirror could be indicative of how they would react to an unfamiliar animal. Studies using primates in mirror self-exploration tests have been criticised for using paralipsis to assume that this means having a concept of self, or some kind of defined self-awareness (Heyes, 1994; but compare Gallup *et al.*, 1995; Heyes, 1995; De Veer & Van Den Bos, 1999). Tests with cotton-top tamarins have suggested that these callitrichids may have at least some ability to recognise themselves in mirrors (Hauser *et al.*, 1995), although this conclusion has been criticised (Anderson & Gallup, 1997, Hauser & Kralik, 1997). As marmosets are relatively olfactory animals, it could be that if an image does not smell like a conspecific, they would not react to it as if it were an intruder. It is also possible, although perhaps less likely, that they do not recognise the image in the mirror as a monkey at all. Most studies using mirrors that are studying self-recognition habituate primates to the mirror's presence for a relatively long period of time (Heyes, 1994). As presentations during the current study were very short (240 seconds), it is highly unlikely that animals are experiencing any self-recognition, even if it is theoretically possible.

6.4.2 Other studies: heritability and methodology

Overall, this study gives little overall support for genetically based variation for personality-like behavioural traits in common marmosets. This is either because such variation truly does not exist, or because the methodology used in the study did not manage to detect it. Why should marmosets have been expected to show heritability in such traits?

The results contrast to those in studies of human personality traits, which consistently display at least some heritability (e.g. Bouchard & Loehlin, 2001). The differences seen could be because marmosets simply do not have heritable behavioural traits in the ways humans do. The response behaviours measured in the marmosets do not directly match those such as extraversion assessed in human studies. Human personality traits defined by psychologists have been done so after intense study and are not simple traits that are measured by novel stimulus presentations (although aspects of them maybe). This study is a first step for assessing marmosets in this way, and due to its simplicity, is not

directly comparable to human personality studies. Having said this, if these varying personality traits and behaviours have underlying essences, based on neurochemistry, that are homologous, some similarities should be apparent. This study does at least demonstrate that behavioural traits of common marmosets can be measured and assessed in this way. It gives hints that with a greater sample size and some refinement of methodology (see below), heritable response traits may be demonstrated, and might in the future be compared more directly to similar traits in humans and across the primates.

Other recent investigations into the heritability of behavioural traits in primates have had some (albeit limited) success in finding genetic influence. In infant rhesus macaques, latency to leave their mother, latency to inspect novel food and exploration and movement during a human intruder test have all been found to have high, significant heritability (Williamson *et al.*, 2003). It is worth noting that in this study, as well as in the current one, factor analysis of measured behaviours did not lead to a set of significant heritable characteristics. Dominance (as a broad personality trait rather than related to actual dominance in a group) was found to be significantly heritable in chimpanzees, but human "Big Five" factors were not (Weiss *et al.*, 2000). Social impulsivity has been found to be heritable in vervet monkeys (Fairbanks et al., 2004).

It could be suggested based on these results alone that common marmosets simply do not have consistent, heritable behavioural traits. The primates in the studies discussed above are all catarrhines. Old World monkeys, humans and other apes all are more closely related to each other than they are to the common marmoset (Purvis, 1995). It is possible that this phylogenetic distance between common marmosets and Old World species may include differences in the genetic basis of behaviours. To my knowledge, there are no other studies of the heritability of behavioural traits in a platyrrhine species, and there are no studies in strepsirhines. Marmosets, along with the other mammals, do share the same neurochemicals as humans, including dopamine, serotonin and other monoamines (Hornung, 1997). Indeed, similarities are enough that researchers are confident in using *the* common marmoset as a model in a wide range of biomedical, psychological and behavioural studies (for discussion see Pryce, 1997). This evidence, when added to knowledge about variation in behavioural traits in marmosets (e.g. Blackwood, 2000), suggests it is counterintuitive to argue against a genetic influence on behavioural traits.

Is there a reason in the experimental methods of tests that may have caused a different outcome? The above studies did not control for multiple comparisons regarding the pedigree of the animals used in each analysis. Even so, the significance of several of the trait heritabilities were below a significance level of 0.05 and this was not the case in the current study (it is worth noting, however, that if many traits were tested, the alpha required for significance would be well below 0.05). Sample size in the current study was limited. As animals were part of a breeding colony, they could only be accessed outside of husbandry times, and on certain days whole rooms of animals were unavailable. Animals were also moved out of the colony on a regular basis, so individuals were only available for testing for a limited time. Other studies have used more animals in their analyses. Fairbanks *et al.*'s (2004) study of social impulsivity included 352 vervet monkeys; Rogers *et al.*'s (2004) investigation into monoamine metabolites, used 271 rhesus macaques; Weiss *et al.*'s (2000) study of personality traits in gorillas had a sample size of 145. One study, of chimpanzees, (Williamson *et al.*, 2003), did have a sample size of 85; only 17 more than the current study. They found heritable differences in behaviours similar to the ones looked at here. Sample sizes can often be higher than this, for instance, a study of flatcoated retrievers noted observations for behavioural traits on between 800-1150 animals. Human twin study figures are often in their thousands or tens of thousands (Plomin *et al.*, 2000). It is thus possible that a low sample size may have caused problems, but not necessarily, as studies with less than 100 animals have found heritable traits before.

6.4.3 The effect of sex on response

In Chapter Four, limited sex differences were seen in response. These tests looked at stimuli individually, and found males were quicker to come into close contact with both of the food-related stimuli (RIT and RIW), and quicker to approach the tube, TUB, and the syringe, SYR. These differences are echoed in the heritability analyses where sex was seen as a significant covariant. For instance, when corrections for multiple comparisons were not imposed, sex was seen to affect latency to close contact with novel stimuli (a group including TUB and SYR). Reactions to food were also affected by sex differences. Sex was a significant co-variant at a 0.05 level for the duration of proximity to the food related stimuli, and at less than a 0.0025 level for latency to close contact with a food-related stimulus. This means then that the sex of an animal is

affecting its reaction to unusually presented foodstuff. As discussed in section 4.4.6 and 4.7, previous experimental studies where females were found to be more responsive (e.g. Box, 1988; Visalberghi *et al.*, 2003), are not supported by this result. It also contrasts a non-primate study, where females were found to be quicker to obtain food from an unfamiliar feeder (guppies, *Poecilia reticulata*; Laland and Reader, 1999). Where food is involved in a novel stimulus or situation, females may exercise priority of access over males (Box, 1997; Box, 1999; Petto & Devin, 1988). This priority of access could be due to factors other than speed of contact with a food item, such as males demurring, or females forcing them to keep away. This is not something that would be seen in the current study, as individuals were tested in isolation.

6.5 Summary: the heritability of behavioural traits

This study has not demonstrated any clear evidence of heritability in response behaviours in this captive colony of common marmosets. There were, however, some non-significant indications of heritability in response to mirrors and to unattractive stimuli. Of the five behavioural measures, only latency to approach showed (even non-significant) heritabilities for all the stimulus groups. This indicates that it may be the best method of assessing variability in behaviour, from the point of view of the effect of genes on variability.

The findings contrast with studies of behavioural traits in other primate species, where a limited number of heritable characteristics have been found. This difference could be either because of problems with the methodology of the study, for instance a relatively small sample size, or due to the fact that marmosets do not have the heritable characteristic traits found in Old World primates.

Chapter Seven

Discussion

7.1 Introduction

Primates and other animals demonstrate individual variation in behaviour. There are numerous possible causes for such variation, including sex, age, genetic differences and physical differences. The response to new objects, environments, and situations is a varying behavioural trait found in a wide range of species (Wilson *et al.*, 1994). Variation in response has the potential to be measured by a relatively simple test paradigm, such as novel stimulus presentation (Blackwood, 2000). This study attempted to assess variation in response behaviours in common marmosets, and investigate the possible underlying causes. The four main hypotheses of the research were:

1. Common marmosets display individual variation in response to novel stimuli that has a measurable genetic influence

2. Sex affects response to novel stimuli in common marmosets, with females being more responsive than males

3. Age affects response to novel stimuli in common marmosets, with older subadults being the most responsive

4. Weight affects response to novel stimuli in common marmosets

The purpose of this final chapter is to review the main findings of this study in relation to these four hypotheses and focus on several of the major issues arising from the research. In doing so, potential avenues for further research will be discussed.

7.2 Summary of results

In this study, responsiveness to novel stimuli in a home cage environment was assessed using five behaviours related to response (visual attendance, latency to first approach,

latency to first contact, duration of proximity and duration of contact). These behaviours were demonstrated to be measurable and variable across the 68 animals tested. The effects of sex, age and weight on response were assessed, in order to investigate the four hypotheses stated above. During the analysis, the nine stimuli used in the testing were grouped together by variation in how animals responded to them. The four categories defined were: mirror, unattractive stimuli, food-related stimuli and novel stimuli. Principal component analysis was then used to create continua of responsiveness for these stimulus groups, by combining the behaviours recorded into more general measures of response. There were either one or two continua per stimulus group. A single continuum described most of the variation seen in response, where there were two continua per group, latency to response and duration of response were separated. In order to investigate any genetic influence on variation in response, heritability analyses were carried out on both these general response continua and the original behavioural measures. The major findings of the study relating to the four hypotheses stated above were:

1. There is no significantly heritable individual variation in response behaviours or continua derived from them

2. There are some limited differences in responsiveness between sexes, especially in response to novel, food-related stimuli. Males are more responsive than females

3. Responsiveness does not vary significantly in accordance with an animal's age

4. Responsiveness does not vary significantly in accordance with an animal's size

None of the four hypotheses tested were supported. There were sex differences in behaviour (Hypothesis 2), but only in a limited sense, and in the opposite direction to that predicted.

7.3 Common marmosets and variation in response behaviours

7.3.1 Responsiveness and sex

Overall in the study there were few instances of sex differences in response. Where differences between the sexes were seen, males were more responsive than females, touching novel food containing stimuli more quickly. This is contrary to previous research in callitrichid species. In tamarin species, when differences in response to food and food related stimuli occur, females are more responsive than males (Box, 1998; Box *et al.*, 1995). In two species of marmoset (common marmosets and black tufted-eared marmosets), females are more responsive to additional food when it is offered (Box & Smith, 1995). In these previous studies, animals were tested in groups, and arguments such as female priority of access can be used to explain differences (Box, 1997). Female common marmosets, however, have also been shown to be more responsive to unfamiliar food tasks when tested in isolation (Yamamoto *et al.*, 2004). This demonstrates a difference in response behaviour between males and females that is not dependent on the presence of a member of the opposite sex.

There are two potential explanations for the differences in results seen between the current study and previous ones. One possibility is that the significant results of previous studies, which use relatively small sample sizes, are due to stochastic variation rather than real differences. For example, Trivers and Willard hypothesised that natural selection might favour an ability to adjust offspring sex ratio depending on environmental conditions (Trivers & Willard, 1973). A number of studies have reported general support for the hypothesis (Clutton-Brock & Iason, 1986; Godfrey &Werren, 1996). In a meta-analysis of the Trivers-Willard hypothesis in primates, Brown & Silk (2002: 11253), however, demonstrated that:

> much of the observed variation in sex ratios of high- and low- ranking females, which often has been interpreted in adaptive terms, actually may be the product of stochastic variation in small samples.

That is, as sample sizes increase, evidence for the effect in primates disappears. In Brown and Silk's study sample sizes reach beyond 1000, and it is at these higher numbers that no effect is seen. The sample size in the current study is greater than many that have contrasting results, but still below 100. It is possible that if other studies

with larger sample sizes are conducted, the evidence for females being more responsive, or in fact any significant differences at all, may disappear.

If males are thus more responsive than females even in a species such as the common marmoset that has little sexual dimorphism, there may be potential consequences for other primates. Differences in behaviour between sexes are often linked to dimorphism (male orang-utans for instance, which are much larger than females, forage more on the ground: Rodman & Mitani, 1987). If sex differences in (non-sexual) behaviours are seen in a species with little dimorphism, it suggests that there may also be innate sex differences in the (non-sexual) behaviour of other primates without accentuated sexual dimorphism, including gibbons and humans.

The second, slightly more complex, explanation for the contrast between the current study and previous callitrichid research is based on differences in the ages and the range of ages of the animals used in different studies. Although age was not shown to have a significant effect on response in the current study, the age range of the animals was limited (10-23 months). All of the female and the majority of the male animals studied by Yamamoto *et al.* (2004) were older than the animals studied here (females 2.5-5 years, males 1.5-5 years). It is possible then that food response behaviours develop and change as animals age; that the slightly higher responsiveness levels of juvenile and sub-adult males in response compared to females alter as females fully develop. What this does not explain is why the male marmosets should be more responsive in the first place, if sex differences only occur as the animals mature. The pattern of males being more responsive than females would be expected for other primate species (especially those with greater sexual dimorphism), based on studies of innovation (Reader & Laland, 2001) and Bateman's rule (Bateman, 1948; Futuyma, 1998). This then suggests that a higher level of responsiveness in males is constant across primate species, and it is at sexual maturity that variations in this occur, caused by (lack of) sexual dimorphism or particular social systems.

In addition to the small differences displayed in response to food related stimuli, males also looked at the tube (TUB) for longer and approached the syringe (SYR) more quickly. Both of these stimuli are cylinders of a relatively similar diameter and length. It is possible then that the shape of the stimuli has an effect on the animals' responsiveness. Marmosets are extractive foragers; they may be reacting to the tubes as

if they are branches from which gum can be extracted, or objects that could contain hidden animal prey (although animals in this study would never have been exposed to live prey). If responses are based on reaction to a possible food source, the same issue as for the food related stimuli is raised: should sex differences again not echo what is seen in previous work, with females being more responsive? Further research using tubes or more natural, 'organic' stimuli could investigate several of the issues raised here; for instance, by examining how the animals interact with the object (holding with hands, gnawing, etc.) to see if they are treating it as a potential food source.

7.3.2 Age

Contrary to the findings of some other studies (Millar *et al.*, 1988; Rogers, 1999), no age differences in response were seen. The lack of age differences could be because of the restricted age range in the group studied (see 7.3.1 for the possible interaction of age and sex factors), or it could be related to the test environment. The relatively heightened responsiveness of young adults could be due to higher motivation or more exploratory behaviour than adults and younger offspring. In other studies of callitrichid responsiveness and exploration that have examined age differences, testing has taken place in more open spaces (e.g., Menzel & Menzel, 1979; McGrew & Mcluckie, 1986), or with other animals present. Adults might have other concerns, such as vigilance, or nursing young offspring. When animals are tested in groups in relatively large areas such differences would be apparent. It is possible that for this test, motivation to explore, as opposed to responsiveness *per se*, was not an issue, as animals were necessarily close to the stimuli on entry to the test cage. With a novel stimulus obviously present and no other animals or opportunities to distract the individual being tested, it is possible that any age differences are reduced.

The lack of age differences over a period like this, juvenile to an adult, suggests that both male and female marmoset behaviour does not change with the onset of sexual maturity. This does, however, contrast with the assertion above that sex differences in the behaviour may develop as animals age, and thus could be used instead in support of the suggestion that differences seen in previous studies between animals of different ages and sex are due to stochastic variation. The only way to resolve these conflicting views would be to study marmoset behaviour through this period of sexual development.

7.3.3 Individual differences in response behaviour based on genetic variation

This study did not demonstrate any clear evidence of heritability in responsiveness, indicating that individual variation based on genetic differences does not have any significant effect on response behaviour in the group of common marmosets studied. It is possible that the results reflect the phylogenetic difference between marmosets and other primates where the effect of genes on such behaviour has been studied (see 6.4.2). Genetic differences in response behaviour, however, have been demonstrated in non-primate species such as mice (Flint *et al.*, 1995; Duluwa *et al.*, 1999), and many of the same neurochemicals that underlie behavioural variation, including dopamine, serotonin and other monoamines are shared across mammal species (Hornung, 1997; Mehlman *et al.*, 1994, 1995; Benjamin *et al.*, 1996). Both intuitively, and based on other evidence, it thus seems unlikely that a total lack of genetic influence on response behaviour would be the case. This study does show, however, that such genetic influences are relatively unimportant compared to sex differences in response behaviour. It would be interesting to see exactly what level of variation marmosets did show in the genes where differences have been demonstrated for humans and other mammals. Preliminary studies looking at variation in genes related to neurochemicals connected to response hint that it may be rather low (De Ruiter, personal communication).

The lack of evidence for genetic variation in the behaviours investigated, together with the lack of positive evidence for human-like personality traits from other studies (e.g., King and Figueredo, 1997), suggests caution in ascribing human personality to non-human primates. Humans and non-human primates may share the neurochemicals that underlie behavioural traits, but they do not necessarily express themselves in the same way at a higher, behavioural, level, as described by human "personality".

7.3.4 Physical variation and differences in response

With the exception of differences between sexes, physical variation (as measured in this case by weight) did not have any effect on the response of the animals tested. This indicates that variation in response in the common marmoset is not part of a behavioural strategy (Wilson *et al.*, 1994; Wilson & Yoshimura, 1994) that is based on physical characteristics of the animals related to size, at least across the age range studied and thus the developmental period from juvenile to young adulthood. This then suggests that

more positive responses to novel stimuli and situations are not, in a captive situation, leading to aminals having greater access to food and thus growing larger than their contemporaries. Extreme cases of size affecting behaviour are often related to sex and sexual strategies, for example in orang-utans (Utami *et al.*, 2002), coho salmon (Gross, 1985), and ruffs (van Rhijin, 1973, 1983). In these species, males have distinct morphs that relate to different sexual strategies. Because of the social systems of marmosets, such differing strategies within males may not be necessary. Unrelated males are known to share one social group (Nievergelt *et al.*, 2002), however, so there is potential for varying male strategies to emerge, albeit at a less obvious level than for those species mentioned above with different morphs. Males within natural social groups might show variation in response behaviour related to strategies that is not seen when animals are in single sex peer group cages.

7.4 Potential for further studies

The results of this study and the above discussion can be used to form several suggestions for future research, and some points have been mentioned therein. The increased responsiveness seen in males opens up the possibility of investigating response and related behaviours in a range species with little or no sexual dimorphism, to see if large sex differences in behaviour are retained. The differences between this and other studies in sex differences in response could be explained if these sex differences change as animals mature. This does, however, contrast with other results from this study showing that behaviours do not alter significantly across age groups. The results of both sex differences and age variation the study thus prompt the need for further study of responsiveness in marmosets at different ages to ascertain why males might be more responsive than females at a younger age. Longitudinal studies of marmosets of both sexes over their lives would show how sexual maturity affects non sexual behaviours, and whether and when these behavioural differences between males and females do occur.

If this study and its methodology were to be followed up and expanded upon, an obvious improvement would be an increase in sample size, meaning that heritability analyses could be more reliable. This could be achieved by expanding testing to several other breeding colonies in the United Kingdom, as long as conditions and husbandry practices would allow for a similar experimental set up. Further specific points to be

addressed when looking at how this study might be directly built upon include: should what was measured be changed; should the stimuli used be changed; and should the individuals studied be changed? These questions are addressed below.

7.4.1 Using behavioural measures

All the behaviours recorded in the study were useful for describing responsiveness, but for investigating heritability of response behaviours, latency to approach was most useful. This behaviour illustrates something of an immediate reaction to the stimulus, the animal's first response to a novel item in the home cage before it has moved in close enough to examine it. Thus it might relate more to an animal's immediate pattern of response than behaviours (such as duration of proximity or contact), which depend more on the quality of the stimulus itself. When this measurement does not demonstrate individual variation across the sample (for example, if no or very few animals approach a stimulus), other behaviours such as visual attendance can be used to illustrate awareness of the stimulus. If only these two behavioural measures are used, however, much information about response could be lost. The two duration scores recorded similar behaviours, as did the two latency scores. The principal component analyses of Chapter Five demonstrated that the scores could either be represented by a single response continuum of a pair, with one describing latency and one describing duration. It would be simpler then in future studies to limit recording to one of the latency scores, latency to approach, and one duration score. If duration of contact was measured, then latency to contact could be calculated based on those observations if required. Even if latency to approach is the most informative behaviour in assessing response, future presentation tests should continue to record the measures used here, or at least visual attendance and one measure of duration in addition to latency scores. Video and sound recording, not possible in this study, would allow for even finer measurements of response. Recording of vocalisations would allow for extra elucidation of the emotional state of the animal, and perhaps how it perceived a particular stimulus (as a threat or conspecific, for example). A finer measure of gaze, combined with vocalisations would help to confirm that it was the stimulus that the animal was responding to and not something else in the environment that the observer was perhaps unaware of.

7.4.2 Developing the stimuli used in presentations

A variety of stimuli were used in this study to see if different aspects of response behaviour could be observed. Variation in responsiveness could be measured for all of the stimuli presented, but the presence of food and potential social influences allow for functionally different responses to be observed. In an extension of the methodology of the current study, a suite of stimuli could be used to investigate a range of different response behaviours. Stimuli that offer more interactive opportunity, such as moving parts, or stimuli that are mobile, such as remote controlled toy cars (Williamson *et al.*, 2003), may elicit more varied responses. One novel stimulus with a simple moving arm or pendulum could be used to assess responsiveness (subject to preliminary testing to ensure variation in reaction for a small sample of animals). As it is difficult to differentiate whether an animal is responding to the colour, shape or texture of an object, this approach does increase uncertainty as to what aspect of the stimulus is driving the response behaviour. In the case of the car, for instance, is it the colour and shape of the stimulus itself, or the movement or noise associated with it, which has the greatest effect on the animal's response? Rather than a food related stimulus, novel fruit could be used to measure response behaviour to food. This would remove the possibility that the individuals might not initially be aware that a food reward is present in an inorganic stimulus, although it may also reduce variability in response, if it was immediately obvious that it was edible.

In the study, a mirror was used as a potential social stimulus. There is much discussion in the literature as to whether different primate species recognise themselves in mirrors, and if this indicates that they have a complex concept of self (see Heyes, 1994, 1995; Gallup *et al.*, 1995; De Veer & Van Den Bos, 1999; for a taste of the discussion). Although it is generally thought that only certain apes can recognise themselves, it has been suggested that cotton-top tamarins also have the ability (Hauser *et al,* 1995; but see Anderson & Gallup, 1997; Hauser & Kralik, 1997). Whatever position is taken, in this study of marmosets, each subject was only exposed to a mirror for a 240-second period. They had not been exposed to mirrors previously and any metalwork in and around their cages was tarnished to be non-reflective. Most studies using mirrors that are studying self-recognition habituate primates to the mirror's presence for a relatively long period of time (Heyes, 1995). Thus in the current study, it is highly unlikely that the subjects

would experience any self-recognition, even if it were theoretically possible after lengthy exposure.

To study social responses specifically, a less equivocal stimulus such as a model or taxidermised conspecific (see Barros *et al.*, 2000; for an example of using taxidermised animals as stimuli), or a live, novel, conspecific in an adjoining cage (Fairbanks, 2001), could be used to assess variation.

Automating response measurements and stimulus presentation

The causes of variation in many behaviours can be investigated using a similar test paradigm to the temporarily isolated home cage presentations used here. During the data collection period of this study, a possible tool for automated recording of response and learning behaviour was researched and developed. The CBS6000 (Custom Biotelemetry Systems) is a stimulus presentation panel that attaches to the home cage of the animal being tested (Figures 7.1A and 7.1B). The stimulus panel can have a variety of presentation tools attached for the animal to interact with, such as lights, touch panels and food or liquid dispensers. The system is controlled by a microcomputer that is pre-programmed by a PC. The microcomputer is in a portable data logger (Figure 7.1B) that attaches to the stimulus panel. The data logger both controls the stimulus presentations on the panel, and records the subject's responses. Data are downloadable onto a spreadsheet or database programme on a PC.

During development, tests were carried out using a stimulus array consisting of a light key and a milkshake dispenser and licker (see Figure 7.1A). The data logger recorded when the light stimulus was on and any touches by the subject; the milkshake dispenser recorded when ever a reward was pumped into the licker, and whenever the animal took a reward, by using an infra-red sensor. Three stimulus presentation phases were programmed into the panel. In the first session the CBS6000 was programmed to give out a milkshake reward independent of any action of the subject animal. This would habituate the marmoset to both the stimulus panel and the milkshake dispenser and licker. The main testing phase involved the presentation of a stimulus; the light key would be illuminated. If the animal responded "correctly", i.e. by pushing the light key, it received a small amount of milkshake (c. 0.1ml) as a reward. After the animal had learned to push the light key for a milkshake reward and had done so successfully over

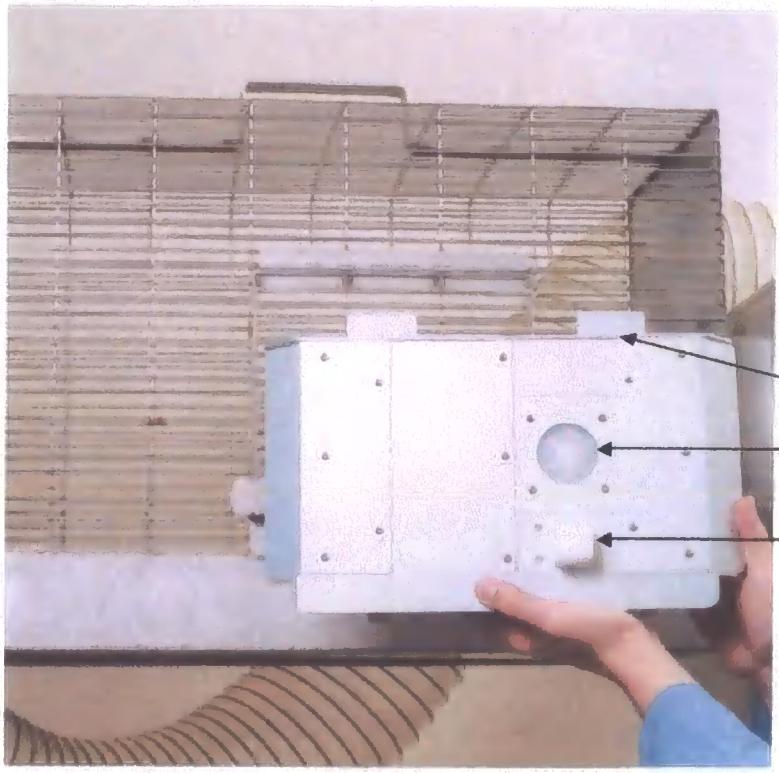


Figure 7.1A: The front view of the CBS6000, showing the light stimulus and milkshake dispenser

Figure 7.1B. The CBS6000 automated response recorder and data logger

several repeats, the third phase was introduced. Here, the light key stimulus was still presented, but no reward was given. The data logger would record for how long an animal persisted to press the light key and lick the milkshake dispenser's licker.

This simple series of presentations used in developing the CBS6000 illustrates its potential to record several types of behavioural response. The first regime gives a measure of how quickly the animal will approach and touch a novel stimulus, and the second regime records how quickly it can learn how to carry out a simple task for a reward. The third regime, where the stimulus is presented with no reward, measures the subject's "reward dependence": how long will it persist in trying to gain a reward that is no longer obtainable. Reward dependence is a personality trait defined, along with novelty seeking, in Cloninger's Tri-Dimensional Personality Questionnaire (Cloninger, 1987). In humans the trait can reflect a need for approval, and remains consistent across age groups (Hamer & Copeland, 1998; Cloninger, 1987). An animal (or person) scoring high for reward dependence would show resistance to the extinction of an unrewarding behaviour. This can be tested in a laboratory situation through "frustrative non-reward" tests such as the third test regime described above. As the presentations were carried out during the development of the system and were preliminary, the sample size of individuals tested was too small for analysis of the behaviours themselves.

This is only one example of how a system such as the CBS6000, which is itself still in development, can be used to measure a range of behavioural traits in a simple, home cage based situation. The stimulus panel is designed so that other stimuli such as mirrors, different light displays or simple cognitive tasks could be added for different presentation tests. The system has the bonus of not requiring an observer to be in the line of sight of the subject, with the potential to distract it. If the observations could be automated, or the observer could at least be outside the vision of the animal being tested (this was not possible in the current study), it may reduce the time needed for habituation and also the risk of void presentation sessions if the animal is more interested in the observer. Automated observations mean that behaviour on a "micro-scale" can be observed, multiple licks of the milkshake dispenser, for instance, that a human observer would not see. It is possible that methods other than direct physical observations on the part of the experimenter can successfully assess responsiveness, but there are drawbacks to using an automated system. An automated stimulus panel could record detail about response behaviour missed by human observation, but it cannot

measure variables such as visual attendance, and latency to approach as opposed to latency to contact the stimulus. Human observers can also take additional note of behaviours that are interesting but not part of the recording brief; an automated system cannot do this.

7.4.3 Choosing study groups for investigating response behaviours

The test groups in this study were chosen for a large part on availability and accessibility. The way the colony is set up, with juveniles and young adults in peer groups rather than family cages, allows for easy testing of temporarily isolated individuals of a certain age. There are some issues raised by this study that would have to be addressed by studying subjects in a different regime. Testing animals of both sexes both individually and in their family groups would address the effect of primacy of access for females. It is possible that the age of animals tested affects sex differences, with females becoming more responsive than males, especially to food, as they get older (see Section 7.3.2). In order to investigate whether this is the case, an experiment would need to test response behaviours of males and females individually across all age groups. If the same individuals were tested within family groups, the differences between primacy of access and responsiveness *per se* could be examined. If, in testing, both food and non food-related stimuli were presented to these individuals, it would be possible to see how reactions differ to them. This could then be extended to either (preferably) a longitudinal study, or wider cross-sectional study of animals of different ages.

7.4.4 Wider implications of variation in responsiveness

Awareness of both the difference between human and non-human primate personality traits (and indeed the difference between non-human primate species) and the individual variation and sex differences that exist within a primate species are important if behavioural models are being used. Monkeys, especially common marmosets, are often used as model organisms in biomedical research that involves behavioural parameters (e.g. Marshall & Ridley, 2003; Ridley *et al.*, 1999; and discussion in Pryce, 1997). When primates are used in models that involve observing the effect of surgery or other invasive measures, it is important that a thorough assessment of the subject's behaviour has been carried out beforehand. If the baseline behaviour is not measured so as to

accurately document the range of behaviours seen in an experimental group, subsequent observations may not accurately describe any change in behaviour. Knowledge of the normal, general behaviour of the animal is required to ensure that behaviours noted during tests are due to the conditions imposed by the experimental condition specifically, rather than a species typical reaction to the experiment in a wider sense. Accounts of marmoset behaviour are thorough (e.g., Stevenson & Poole, 1976), but are often based on relatively small sample sizes. The results of the current study can help to classify the range of response behaviours seen across a population of marmosets. "Unusual" behaviours seen during experiments may be at the tails of a normal distribution of response, rather than an abhorrent movement away from a species standard. In short, studies such as this give valuable information on the range of response seen within a species (indeed a population) as well as giving an idea of species general behavioural responses that are essential for the in-depth understanding of behavioural responses in biomedical experimental work.

In addition to large scale research being used to enhance knowledge of individual variation, short tests of response behaviours could be used to select animals for experimental biomedical work that includes behavioural parameters. A series of short novel stimulus presentations could assess whether an animal is relatively positive in its response to stimuli, or unwilling to approach or investigate novel objects. If an animal is unwilling to interact with novel stimuli and nervous in novel environments, it could be excluded from a research program that required it to be exposed to new experiences. This would avoid investing time and money in the animals training if that individual was to be unusable in the study.

7.5 Conclusions

This study showed that response behaviours are not significantly heritable in the common marmoset group tested, and sex is a more important determinant of response that individual genetic differences, age or weight. The lack of significant heritability in responsiveness illustrates that just because species have a similar underlying biology, that biology may not manifest itself in the same way, especially when it comes to complex behavioural variation, or what is referred to in humans as personality. Responsiveness and variation in response behaviours can be complex to study, both from the point of view the patterns of responsiveness that might be found within a

species, and of the mechanics of the testing itself (for example, what behaviours best illustrate how an animal is responding). These results, when taken with those of other studies, demonstrate the varying effect of sex and age on response to food related novelty, suggesting that as marmosets age, differences between the sexes expand, or even reverse, with females becoming more responsive than males, when the opposite is true at a younger age. It is also possible, however, that the contrast in results with other studies is due to variation in sample sizes, and that as studies use larger numbers of marmosets, patterns of results more similar to the current study will emerge.

Appendix One

Subject information, including sex, age at testing, birth weight and test weight, and litter size. The animal's code refers to its sex and its age relative to the other subjects; lower numbers indicate older animals. Group gives the group number of the animal, groups are numbered in the order tested. Group 7 was split into two during testing due to aggression. Birth weight was taken from colony records; test weight was taken from the monthly weighing closest to the beginning of habituation for that animal's group. Age of testing was calculated using the animal's date of birth from colony records and the first day of habituation for the group.

Code F= Female M=Male	Group	Birth weight (g)	Test weight (g)	litter size	age at testing (days)	Age at testing (months and days)
12F	1	30.2	299	3	399	13m4d
13F	1	27.3	316	3	399	13m4d
16F	1	29.4	310	2	385	12m20d
19F	1	30.6	294	2	373	12m8d
20F	1	30.4	287	4	369	12m4d
21F	1	33.1	337	4	369	12m4d
23F	1	34.2	300	2	361	11m26d
24F	1	33.5	275	2	361	11m26d
26F	1	36.7	292	2	348	11m13d
27F	1	33.7	329	2	348	11m13d
28F	1	32.1	300	2	339	11m5d
29F	1	34.4	271	2	337	11m3d
31F	1	37	317	3	322	10m18d
11M	2	28.3	316	3	494	16m9d
14M	2	37.9	350	2	476	15m19d
15M	2	36.7	350	2	476	15m19d
17M	2	30	307	2	473	15m16d
18M	2	38.7	363	1	462	15m7d
22M	2	35.5	378	4	457	15m2d
25M	2	30.6	326	2	444	14m17d
30M	2	36.8	329	2	425	14m1d
32M	2	36.8	318	3	410	13m14d
33M	2	32.2	300	2	408	13m12d
34M	2	32.5	348	2	408	13m12d
1F	3	38.5	413	2	673	22m4d
2F	3	34.7	473	2	671	22m2d
3F	3	31.6	335	3	665	21m26d
4F	3	32.6	353	3	652	21m13d
5F	3	37.1	359	2	641	21m3d
6F	3	37.9	340	2	612	20m5d
7F	3	37.9	395	2	594	19m17d
8F	3	39.6	397	2	594	19m17d
9F	3	27.8	335	3	578	19m1d
10F	3	32	392	3	569	18m22d
46M	4	30.3	298	3	394	12m29d
47M	4	36.5	330	2	385	12m20d

48M	4	28.2	290	3	379	12m14d
49M	4	31.7	313	3	379	12m14d
50M	4	30.8	322	3	369	12m4d
51M	4	34.3	282	3	348	11m13d
52M	4	29.9	278	3	327	10m23d
53M	4	29.8	278	3	325	10m21d
38F	5a	32.5	347	3	542	17m27d
39F	5a	29.4	330	2	534	17m16d
40F	5a	27.4	374	3	530	17m12d
41F	5b	30.9	363	3	515	16m28d
42F	5b	30.4	338	3	506	16m19d
43F	5b	26.7	334	3	472	15m15d
45F	5b	32	345	3	459	15m2d
56M	6	27.2	334	3	419	13m23d
59M	6	32.4	336	3	410	13m14d
61M	6	27.5	312	3	400	13m5d
63M	6	29.8	290	3	392	12m27d
67M	6	35.5	346	2	379	12m14d
68M	6	36.6	334	2	379	12m14d
69M	6	37.3	364	2	353	11m18d
54F	7	33.9	330	2	372	12m7d
55F	7	31	391	2	448	14m22d
57F	7	29.4	361	2	448	14m22d
58F	7	33.7	402	2	442	14m16d
69F	7	33.1	431	2	442	14m16d
62F	7	31.9	383	3	433	14m7d
64F	7	30.2	318	3	428	14m2d
65F	7	30.5	334	3	420	13m25d
66F	7	28.9	274	2	409	13m14d
70F	7	29.7	288	2	409	13m14d
71F	7	33.3	332	2	381	12m16d
72F	7	37.7	335	2	372	12m7d
Means		**32.6**	**334.1**	**2.5**	**442.5**	

Appendix Two

Behavioural observation check sheet. Whether or not the animal's head was oriented toward the stimulus was noted every ten seconds as √ (yes) or × (no). Behaviours and the time they occurred were entered in the central box area of the sheet; behaviours were given two letter codes for brevity of use. No filled in examples can be given as paper was not allowed to be removed from the colony.

Behavioural Observation Sheet Stimulus_____
Individual _____ Date _____ Time _____

Time (s)	Orient	Behaviours (code and time/s)
10		
20		
30		
40		
50		
1.00		
1.10		
1.20		
1.30		
1.40		
1.50		
2.00		
2.10		
2.20		
2.30		
2.40		
2.50		
3.00		
3.10		
3.20		
3.30		
3.40		
3.50		
4.00		

References

Abbot, D. H., Barrett, J., & George, L. M. (1993). Comparative aspect of the social suppression of reproduction in female marmosets and tamarins. In Rylands, A. B. (Ed), *Marmosets and Tamarins: Systematics, Behaviour, and Ecology*. New York, Oxford University Press.

Alexander, G.M., & Hines, M. (2002) Sex differences in response to children's toys in non-human primates (*cercopithecus aethiops sabaeus*). *Evolution and Human Behavior* **23**: 467-479.

Almasy, L. & Blangero, J. (1998). Multipoint quantitative-trait linkages analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198-1211.

Altman, D.G. (1991). *Practical Statistics For Medical Research.* London: Chapman & Hall.

Anderson J. R. & Gallup, G. G. (1997). Self recognition in *saguinus*? A critical essay. *Anim. Behav.* **54**, 1563-1567.

Bardo, M. T., Donohew, R. L. & Harrington N. G. (1996). Psychobiology of novelty seeking and drug seeking behavior. *Behav Brain Res* **77**, 23-43.

Barros, M., Boere, V., Huston, J.P. & Tomaz C. (2000). Measuring fear and anxiety in the marmoset (*Callithrix penicillata)* with a novel predator confrontation model: Effects of diazepam. *Behavioural Brain Research*, **108** (2): 205-211.

Bateman, A. J. (1948). Intra-sexual selection in Drosophila. *Heredity*, **2**, 349-368.

Benjamin, J., Li, L., Patterson, C., Greenberg, B.D., Murphy, D.L. & Hamer, D.H. (1996). Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nature Genetics* **12**, 81-4.

Blackwood, N.S. (2000). Measuring simple behaviours as a basis for assessing consistent individual differences in common marmosets (*Callithrix jacchus*). University of Durham. Unpublished MSc. Thesis.

Bolig, R., Price, C. S., O'Neill, P.L. & Suomi, S. J. (1992). Subjective assessment of reactivity level and personality traits of rhesus monkeys. *International Journal of Primatology* **13,** 287-306.

Bouchard, T.J. (1994). Genes, environment and personality. *Science* **264**, 1700-1701.

Bouchard, T.J. & Loehlin, J.C. (2001). Genes, evolution and personality. *Behavior Genetics,* **31**, 243-273.

Box, H.O. (1988). Behavioural responses to environmental change. Observations on captive marmosets and tamarins (*Callitrichidae*). *Animal Technology* **39**, 9-16.

Box, H.O. (1997). Foraging strategies among male and female marmosets and tamarins (Callitrichidae): new perspectives in an underexplored area. *Folia Primatol*, **68**, 296-306.

Box, H.O. (2003). Characteristics and propensities of marmosets and tamarins: Implications for studies of innovation. In: Reader, S.M. & Laland, K.N. (eds) *Animal Innovation* pp 197-222. Oxford, Oxford University Press.

Box H.O. & Hubrecht, R.C. (1987). Long-term data on the reproduction and maintenance of a colony of common marmosets (*Callithrix jacchus jacchus*) 1972-1983. *Laboratory Animals*, **21**, 249-260

Box H.O., Roehrhuber, B. & Smith, P. (1995). Female tamarins (*Saguinus* - Callitrichidae) feed more successfully than males in unfamiliar foraging tasks. *Behavioural Processes*. **34**, 3-11

Box, H.O. & Smith, P. (1995).Age and gender differences in response to food enrichment in family groups of captive marmosets (*Callithrix*–Callitrichidae). Anim. Technol. **46**, 11–18

Box, H.O., Yamamoto, M.E. & Lopes, F. A. (1999). Gender differences in marmosets and tamarins: responses to food tasks. *International Journal of Comparative Psychology* **12**, 59-70.

Box H.O. & Yamamoto, M. E. (2002) *Abstracts for the XIXth Congress of the IPS.* Beijing: Mammological Society of China.

Brown, G. R. & Silk, J. B. (2002). Reconsidering the null hypothesis: Is maternal rank associated with birth sex ratios in primate groups? *Proceedings of the National Academy of Sciences,* **99**: 11252-11255.

Buchner, A., Erdfelder, E., & Faul, F. (1997). How to Use G*Power [WWW document]. URL http://www.psycho.uni-duesseldorf.de/aap/projects/gpower /how_to_use_gpower.html

Buirski, P., Kellerman, H. Plutchik, R., Wininger, R. & Buirski, N. (1973) A field study of emotions, dominance and social behavior in a group of baboons (*Papio anubis), Primates* **14**, 67-78.

Buirski, P., Plutchik, R. & Kellerman, H. (1978). Sex differences, dominance and personality in the chimpanzee. *Animal Behaviour* **26**,123-129.

Buirski, P. & Plutchik, R. (1991). Measurement of deviant behavior in a Gombe chimpanzee: relation to later behaviour. *Primates*, **32**, 207-211.

Caine, N.G. & Mundy, N.I. (2000). Demonstration of a foraging advantage for trichromatic marmosets (*Callithrix geoffroyi*) dependent on food colour. *Proceedings of the Royal Society of London B.* **267** 439-444.

Capitanio, J. P. (1999). Personality dimensions in adult rhesus macaques: prediction of behaviours across time and situation. *American Journal of Primatology* **47**, 299-320.

Carey, G. & DiLalla, D. L. (1994). Personality and psychopathology, genetic perspectives. Special issue Personality and Psychopathology. *Journal of Abnormal Psychology* **103**, 32-43.

Cases, O., Seif, I., Grimsby, J., Gaspar, P., Chen, K., Pournin, S., Muller, U., Auet, M., Babinet, C., Shih, J.C. & De Maeyer, E. (1995). Aggressive behavior and altered amounts of brain serotonin and norepinephrene in mice lacking MAOA. *Science* **268**, 1763-1766.

Chamove, A.S., Eysenk, H.J. & Harlow, H.F. (1972). Personality in monkeys: factor analysis of rhesus social behaviour. *Quarterly Journal of Experimental Psychology* **24**, 496-504.

Clarke, A. S.& Boinski, S. (1995). Temperament in non-human primates. *American Journal of Primatology* **37**,103-125.

Clarke, A.S., Kammerer, C. M., George, K.P., Kupfer, D.J, McKinney, W.T., Spence, A. & Kraemer, G.W. (1995). Evidence for heritability of biogenic amine levels in the cerebrospinal fluid of rhesus monkeys. *Biol. Psychiatry* **38**, 572-577.

Clayton, G.A., & Robertson, A. (1957). A experimental check on quantitiative genetical theory. II. The long term effects of selection. J. *Gene.t* **55**: 152–170.

Cliff, N. (1987). *Analyzing Multivariate Data.* San Diego, California, Harcourt Brace Jovanovich.

Cloninger, C.R. (1987). A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry* **44**, 573-588.

Cloninger, C.R., Svrakic, D.M. & Przybeck, T.R. (1993). A psychobiological model of temperament and character. *Arch. Gen. Psychiatry* **50**, 975-990.

Clutton-Brock, T.H. & Iason, G.R. (1986). Sex ratio variation in mammals. *Q. Rev. Biol.*, **61**, 339-373.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, **112**, 155-159.

Corey, D.T. (1978). The determinants of exploration and nephobia. *Neuroscience and Biohavioural Reviews,* **2**, 235-253.

Darwin, C. (1871). *The Descent of Man, and Selection in relation to Sex.* John Murray, London (in Bouchard and Loehlin, 2001).

Day, R.L., Coe, R.E., Kendal, J.K. & Laland, K.N. (2003). Neophilia, innovation and social learning: a study of Intergeneric differences in callitrichid monkeys. *Anim. Behav.* **65**, 559-571.

De Veer, M. W. & Van Den Bos, R. (1999). A critical review of methodology and interpretation of mirror self recognition research in non-human primates. *Anim. Behav.*, **58**, 459-468.

Dixson, A.F. (1998). *Primate Sexuality* Oxford University Press, Oxford.

Drea C.M., (1998). Social context affects how rhesus monkeys explore their environment. *Am. J. Primatol.*, **44**, 205-214.

Dulawa, S. D., Grandy, D. K., Low, M. J., Paulus, M. P. & Geyer, M. A. (1999). Dopamine D4 receptor-knock-out mice exhibit reduced exploration of novel stimuli. *The Journal of Neuroscience* **19,** 9550-9556.

Dyke, B. (1989). PEDSYS, a pedigree data management system users manual. *Population Genetics Laboratory Report No. 3.* San Antonio, Southwest foundation for Biomedical Research.

Ebstein, P.R., Novick, O., Umansky, R., Priel, B., Osher, Y., Blaine, D., Bennett, E.R., Nemanov, L. Katz, M & Belmaker, R.H. (1995). Dopamine D4 receptor (D4D4) exon III polymorphism associated with the human personality trait of novelty seeking. *Nature Genetics* **12**, 78-80.

Emsley, A., Dickerson, G.E. & Kashyap, K.S. (1977). Genetic parameters in progeny-test selection for field performance of strain cross-layers. *Poult. Sci.*, **56**, 121-146.

Epple, G., Belcher, A.M., Küderling, I., Zeller, U., Scolnick, L., Greenfield, K.L., & Smith III, A.B. (1993). Making sense out of scents: species differences in scent glajnds, scent marking behaviour, ans scent copmosition in the Callitrichidae. In Rylands, A.B. (Ed) *Marmosets and Tamarins: Systematics, Behaviour, and Ecology.* Oxford, Oxford University Press.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, **28**, 1-11.

Eysenck, H. J. (1990). Genetic and environmental contributions to individual differences: three major dimensions of personality. *Journal of personality* **58**, 245-261.

Fagen, R. M. (1981). *Animal Play Behavior.* Oxford, Oxford University Press.

Fairbanks, L. A. (2001). Individual differences in response to a stranger: Social Impulsivity as a dimension of temperament in vervet monkeys (*Cercopithecus aethiops sabaeus*). *Journal of Comparative Psychology* **115**, 22-28.

Fairbanks, L.A., Newman, T.K., Bailey, J.N., Jorgensen, M.J., Breidenthal, S.E., Ophoff, R.A., Comuzzie, A.G., Martin, L.J. & Rogers, J. (2004).Genetic contributions to social impulsivity and aggressiveness in vervet monkeys. *Biological Psychiatry* **55**, 642-647.

Falconer, D.S. & Mackay, T.F.C. (1996). *Introduction to Quantitative Genetics* (4th Ed.). Harlow, Prentice Hall.

Faulkes, C. G., Arruda, M. F. & Montiero da Cruz, A. O. (2003). Matrilineal genetic structure within and among populations of the cooperatively breeding common marmoset, *Callithrix jacchus. Mol. Ecol.*, **12**, 1101-1108.

Feaver, J., Mendl, M. & Bateson, P. (1986). A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour* **34**, 1016-1025.

Ferrari, S.F. & Digby, L.J. (1996). Wild *Callithrix* groups: stable extended families*? Am. J. Primatol.* **38**, 19-28.

Ferrari, S.F. & Lopes Ferrari, M.A. (1989). A re-evaluation of the social organisation of the Callitrichidae, with reference to the ecological differences between genera. *Folia Primatologica* **52**, 132-147.

Ferrari S. & Martins, E.S. (1992) Gummivory and gut morphology in two sympatric callitrichids (*Callithrix emiliae* and *Saguinus fuscicollis weddelli*) from Western Brazilian Amazonia. *American Journal of Physical Anthropology*, **88**, 97-103.

Flint, J., Corley, R., Defries, J., Fulker, D.W., Gray, J.A., Miller, S. & Collins, A.C. (1995). A simple genetic basis for a complex psychological trait in laboratory mice. *Science* **269**, 1432-1435.

Forster, F.C. (1995). Exploratory-behavior and learning in laboratory marmosets (*Callithrix jacchus jacchus*) – comparisons between experimental-cage and home-cage activity. *Primates*, **36**, 501-514

Fragaszy, D.M. & Mason, W.A. (1978). Response to novelty in *Saimiri* and *Callicebus*: influence of social context. *Primates*, **19**, 311-331.

Fragaszy, D.M., Visalberghi, E., & Galloway, A. (1997). Infant tufted capuchin monkey's behaviour with novel foods: Opportunism, not selectivity. *Animal Behaviour* **53**, 1337-1343.

French, J.A. (1987). Reproductive suppression in marmosets and tamarins: absence of social effects in the lion tamarin. *Am. J. Primatol.*, **12**, 342.

French, J. A., Abbott, D.H. & Snowdon, C.T. (1984). The effects of social environment on oestrogen excretion, scentmarking and sociosexual behaviour in tamarins (*Saguinus oedipus*). *Am. J. Primatol.*, **6**, 156-67.

French, J. & Inglett B.J. (1991). Responses to novel social stimuli in callitrichid monkeys: a comparative perspective In: Box, H.O. (ed) *Primate Responses to Environmental Change* pp 275-294. London, Chapman & Hall.

Futuyma, D. (1998). *Evolutionary Biology*. Sinauer.

Gallup, G. G. Povinelli. D. J. Suarez, S. D., Anderson, J. R., Lethmate, J. & Menzel, E. W. (1995). Further reflections on self-recognition in primates. *Anim. Behav.* 50, 1525-1532.

Godfrey, H.C.J. & Werren, J.H. (1996). Recent developments in sex ratio studies. *TREE*, **57**, 59-63.

Gosling, S. D. & John, O. P. (1999). Personality dimensions in nonhuman animals: A cross-species review. *Current directions in Psychological Science*, **8**, 69-75.

Greenberg, R.M. (2003) The role of neophobia and neophilia in the development of innovative behaviour of birds. In: Reader, S.M. & Laland, K.N. (eds) *Animal Innovation* pp175-197 Oxford, Oxford University Press.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C & Gelbert, W.M. (1996). *An Introduction to Genetic Analysis.* New York, W.H. Freeman & Company.

Gross, M.R. (1985). Disruptive selection for alternative life histories in salmon. *Nature* **313**, 47-48.

Haig, D. (1999). What is a marmoset? *American Journal of Primatology* **49,** 285-296.

Hailman, J.P. (1962). Pecking of laughing gull chicks at models of the parental head. *Auk* **79,** 89-98.

Hall, S. L. & Bradshaw, J. W. S. (1998). The influence of hunger on object play by adult domestic cats. *Applied Anim. Behav. Sci.* **58**, 143-150.

Hamer, D. & Copeland, P. (1998). *Living With Our Genes*. London: Macmillan.

Hardie, S.M. & Buchanan-Smith, H. (2000). Response of captive single- and mixed-species groups of *Saguinus* to novel nonthreatening objects. *International Journal of Primatology* **21,** 629-648.

Hauser, M.D., Kralik, J., Botto-Mahan, C., Garrett, M. & Oser, J. (1995). Self-recognition in primates; Phylogeny and the salience of species-typical features. *Proc. Nat. Acad. Sci. USA*, **92**, 10811-10814.

Hauser, M. D. & Kralik, J. (1997). Life beyond the mirror: a reply to Anderson & Gallup. *Anim. Behav.* **54**, 1568-1571.

Hearn, J.P. (1987). Marmosets and Tamarins. In: Poole, T.B. (ed) *The UFAW Handbook on the Care and Management of Laboratory Animals* pp568-581. 6th Edn. Essex, England. Longman Science and Technical.

Hershkovitz, P. (1977). *Living New World monkeys (Platyrrhini): with an introduction to primates. Vol. 1.* The University of Chicago Press, Chicago.

Hewitt, G.M. & Butlin, R.K. (1997). Causes and consequences in population structure. In: Krebs, J.R & Davies, N.B. (eds) , *Behavioural Ecology: an Evolutionary Approach.* Blackwell Science, Cambridge.

Heyes, C.M. (1994). Reflections on self-recognition in primates. *Anim. Behav.* **47**, 909-919.

Heyes, C.M. (1995). Self-recognition in primates: further reflections create a hall of mirrors. *Anim. Behav.*, **50**, 1533-1542.

Higley, J.D. Mehlman, P.T. & Higley, S.B. (1992). Cerebrospinal fluid monoamine and adrenal correlates of aggression in free-ranging rhesus monkeys. *Archives of General Psychiatry* **49**, 436-441.

Higley, Mehlman, P.T., Higley, S. B., Vickers, J., Lindell, S.L., Taub, D. M., Suomi, S.J. Linnoila, M. (1996). Excessive mortality in young free-ranging non-human primates with low cerebrospinal fluid 5-Hydroxyindolacetic acid concentrations. *Archives of General Psychiatry* **53**, 537-543.

Hinde, R.A. (1976). Interactions, relationships and social structure. *Man (N.S.)* **11**,1-17.

Hopkins, W.D. & Bennett, A.J. (1994). Handedness and approach-avoidance behaviout in chimpanzees *(Pan). J. Exp. Psych.*, **20**, 413-418.

Hornung, J-P. (1997). Neurotransmission in marmosets. In Pryce, C., Scott, L. & Schnell, C. *Marmosets and Tamarins in Biological and Biomedical Research.* Salisbury: DSSD Imagery.

Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Wadsworth.

Inoue-Murayama, M., Niimi, Y. Takenaka, O. & Murayama, Y. (2000). Allelic variation of the dopamine receptor D4 gene polymorphic region in gibbons. *Primates*, **41**, 383-392.

Jackson, J.E. (1991). *A User's Guide to Principal Components*. New York, Wiley.

Jacobs, G.H. (1998). A perspective on colour vision in platyrrhine monkeys. *Vision Research* **38,** 3307-3313.

Jang K.L., McCrae, R.R, Angleitner, A., Riemann, R. & Livesley, W.J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology*, **74**, 1556-1565

Jordão, J. M. & Oliveira, R. F. (2001). Sex differences in predator evasion in the fiddler crab *Uci tangeri* (Decapoda: ocypodidae). *Journal of Crustacean Biology*, **21**, 948-953.

Joubert, A & Vauclair, J. (1986). Reaction to novel objects in a troop of Guinea baboons: approach and manipulation. *Behaviour* **96**, 92-104.

Kagan, J., Resnick, J. S. & Snidman, N. (1988). Biological basis of childhood shyness. *Science* **240**, 167-171.

Kagan, J. & Snidman, N. (1991). Infant predictors of inhibited and uninhibited profiles. *Psychological Science* **2**, 40-44.

Kellerman, H. (1966). The emotional behaviour of dolphins (*Tursiops truncatus)*: Implications for psychoanalysis. *Int. Ment. Health Letters,* **8,** 1-7.

King, J.E. & Figueredo, A.J. (1997). The five-factor model plus dominance in chimpanzee personality. *J. Res. Person.* **31**, 257-271.

Koenig, A. (1995). Group size, composition, and reproductive success in common marmosets (*Callithrix jacchus*). American journal of Primatology **35**, 311-317.

Laland. K. N. & Reader, S. M. (1999). Foraging innovation in the guppy. *Anim. Behav.* **57**, 331-340.

Lefebvre, L. (2000). Feeding innovations and their cultural transmission in bird populations. In *The Evolution of Cognition.* Heyes, C.F. & Huber, L. (Eds). London: MIT Press.

Lindberg, S. Strandberg, E. & Swenson, L. (2004). Genetic analysis of hunting behaviour in Swedish flatcoated retrievers. *Applied Anim. Behav. Sci.* doi:10.1016/j.applanim.2004.03.07

Line S.W. & Morgan K.N. (1991). The effects of two novel objects on the behavior of singly caged adult rhesus macaques. *Laboratory Animal Science,* **41**, 365-369.

Maestriperi, D. Wallen, K. & Carol, K. A. (1997). Infant abuse runs in families in group-living rhesus macaques. *Child Abuse and Neglect* **21**, 465-471.

Majolo, B., Buchanan-Smith H.M., Bell, J. (2003). Response to novel objects and foraging tasks by common marmoset (*Callithrix jacchus*) female pairs. *Lab animal*, **32**, 32-38.

Marshall, J. W. & Ridley, R. M. (2003). Assessment of cognitive and motor deficits in a marmoset model of stroke. *ILAR J.***44**, 153-60.

Martau, P. A. Caine, N. G. & Candland, D. K. (1985). Reliability of the emotions profile index, primate form, with *Papio hamadryas, Macaca fuscata,* and two *Saimiri* species. *Primates,* **26**, 501-505.

Martin, P. & Bateson, P (1993). *Measuring Behaviour. An Introductory Guide* (2nd edn.). Cambridge University Press, Cambridge.

Mather, J.A. & Anderson, R.C. (1993). Personalities of octopuses (*Octopus rubescens*). *J. Comp. Psych.*, **107**, 336-340.

McGrew, W.C. & McLuckie, E.C. (1986) Philopatry and dispersion in the cotton-top tamarin (*Saguinus o. oedipus*): an attempted laboratory simulation. *Int. J. Primatol.*, **7**, 401-22.

McGuire, M.T. Raleigh, M. J. & Pollack, D. B. (1994). Personality features in vervet monkeys- the effects of sex, age, social status and group composition. *American Journal of Primatology*, **33**, 1-13.

McRae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *J. Pers.*, **60**, 175-215.

Menzel , E. W. & Menzel, C. R. (1979). Cognitive, developmental and social aspects of responsiveness to novel objects in a family group of marmosets (*Saguinus fuscicollis). Behaviour* **70**, 251-279.

Mehlman, P.T., Higley, J.D., Faucher, I., Lilly, A.A., Taub, D.M., Vickers, J, Suomi, S.J. & Linnoila, M. (1994). Low CSF 5-HIAA concentrations and severe aggression and impaired impulse control in non-human primates. *American Journal of Psychiatry* **151**, 1485-1491.

Mehlman, P.T., Higley, J.D., Faucher, I., Lilly, A.A., Taub, D.M., Vickers, J, Suomi, S.J. & Linnoila, M. (1995). Correlation of CSF 5-HIAA concentration with sociality and the timing of emigration in free-ranging primates. *American Journal of Psychiatry* **152**, 907-913.

Millar, S. K., Evans, S. & Chamove, A.S. (1988). Older offspring contact novel objects soonest in Callitrichid families. *Biology of Behaviour*, **13**, 82-96.

Milligan, S.R., Sales, G.D. & Khirnykh, K. (1993). Sound levels in rooms housing laboratory animals: an uncontrolled, daily variable. *Physiology and behaviour* **53**, 1067-1076.

Monany, V. (2000). *Animal Experimentation: a guide to the issue.* Cambridge, Cambridge University Press.

Mousseau, T.A. & Roff, D.A. (1987). Natural selection and the heritability of fitness components. *Heredity,* **59**, 181-97.

Nievergelt, C.M., Digby, L.J., Ramakrishnan, U. & Woodruff, D. S. (2002). Genetic analysis of group composition and breeding system in a wild common marmoset (*Callithrix jacchus*) population. *Int. J. Primatol.*, **21**, 1-20.

Nievergelt, C.M. &Martin, R.D. (1999). Energy intake during reproduction in captive common marmosets (*Callithrix jacchus*). *Physiology & Behavior,* **65**, 849-854.

Petto, A.J. & Devin, M. (1988). Food choices in captive common marmosets (*Callithrix jacchus*). *Laboratory Primate Newsletter*, **27,** 7-9.

Platt, M.L, Brannon, E.M. Briese, T.L. &. French, J.A. (1996). Differences in feeding ecology predict differences in performance between golden lion tamarins (*Leontopithecus rosalia*) and Wied's marmosets (*Callithrix Kuhli*) on spatial and visual memory tasks. *Anim. Learn. Behav.* **24**, 384-393

Plomin, R., DeFries, J.C., McClearn. G.E. & McGuffin, P. (2000). *Behavioral Genetics* (4th Edn.). New York, Worth Publishers.

Plutchik, R. (1965). What is an emotion? *J. Psychol., **61,*** 295-303.

Plutchik, R. (1970). Emotions, evolution and adaptive processes. In: *Feelings and Emotions: The Loyola Symposium.* Arnold, M. (Ed.), Academic Press, New York.

Poole, T.B. (1990). Environmental enrichment for marmosets. *Animal Technology* **41**, (No. 2); http:www.awionline.org/Lab_animals/biblio/at-poole.html

Poole, T.B. & Evans, R.G. (1982). Reproduction, infant survival and productivity of a colony of common marmosets. *Primate Report*, **22**, 3-12

Purvis, A. (1995). A composite estimate of primate phylogeny. *Phil. Trans. R. Soc. Lond. B*, **348**, 405-421.

Pryce. C. R. (1997). Evolutionary and comparative biology: their significance for callitrichid research. In Pryce, C., Scott, L. & Schnell, C. *Marmosets and Tamarins in Biological and Biomedical Research.* Salisbury: DSSD Imagery.

Quinn, G. P. & Keough, M. J. (2002). *Experimental Design and Data analysis for Biologists.* Cambridge: Cambridge University Press.

Reader S.M. & Laland, K.N. (2001). Primate innovation: sex, age and social rank differences. *Int. J. Primatol.*, **22**, 787-806.

Ridley, R. M., Pugh, P., MacLean, C. J. & Baker, H. F. (1999). Severe learning impairment caused by combined immunotoxic lesion of the cholinergic projections to the cortex and hippocampus in monkeys. *Brain Res.* **836,** 120-38.

Roberts, D.F., Billewicz, W.Z. & McGregor, I.A. (1978). Heritability of stature in a West African population. *Ann. Hum. Genet.* **42**, 15-24.

Robertson, F.W. (1957). Studies in quantitative inheritance. XI. Genetic and environmental correlation between body size and egg production in *Drosophila melanogaster. J. Genet.* **55**,428–443

Rodman, P.S. & Mitani, J.C. (1987). Orangutans: sexual dimorphism in a solitary species. In: Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W. & Struhsaker, T. T. (eds) *Primate Societies* pp. 146-154. Chicago, University of Chicago Press.

Roff, D.A. & Mousseau, T.A. (1987). Quantitative genetics and fitness: lessons learned from *Drosophila. Heredity*, **58**, 103-18.

Rogers, J., Brent, L., Martin, L.J., Comuzzie, A.G. & Kaplan, J. (2002). Genetics and primate social behavior: a genetic analysis of behavioural responses to a social stimulus among baboons. *Abstracts for the XIXth Congress of the IPS.* Beijing: Mammological Society of China.

Rogers, J., Martin, L.J., Comuzzie, A.G., Mann, J.J., Manuck, S.B., Leland, M. & Kaplan, J.R. (2004). Genetics of monoamine metabolites in baboons: Overlapping sets of genes influence levels of 5-hydroxyindolacetic acid, 3-hydroxy-4-methoxyphenylglycol, and homovanillic acid. *Biol. Psychiatry* **55**, 739-744.

Rogers, L.J. (1999). Factors associated with exploration in marmosets: age, gender and hand preferences. *International Journal of Comparative Psychology,* **12**, 93-109.

Rothe, H. & Darms, K. (1993). The social organization of marmosets: a critical evaluation of concepts. In Rylands, A.B. (Ed) (1993) *Marmosets and Tamarins: Systematics, Behaviour, and Ecology* 176-199. Oxford, Oxford University Press.

Rowe, N. (1996). *The Pictorial Guide To Living Primates.* New York: Permagon Press.

Rutledge, J.J., Eisen, E.J. & Legates, J.E. (1973). An experimental evaluation of genetic correlation. *Genetics*, **75**, 709-726.

Rylands, A.B. (1989). Sympatric Brazilian callitrichids: the black tufted-ear marmoset, *Callithrix kuhli*, and the golden-headed lion tamarin, *Leontopithecus chrysomelas*, *Journal of Human Evolution* **18**, 679-695

Rylands, A.B. (Ed) (1993) Marmosets and Tamarins: Systematics, Behaviour, and Ecology. Oxford, Oxford University Press.

Sambrook, T.D. & Buchanan-Smith, H.M. (1997). Control and complexity in novel object enrichment. *Animal welfare,* **6,** 207-216.

Schneider, M.L., Moore C.F., Suomi, S.J., Champoux, M., (1991). Laboratory assessment and environmental enrichment in rhesus-monkey infants (*Macaca mulatta*). Am. J. Primatol. **25**, 137-155.

Scott, L. (1991). Environmental enrichment for single housed common marmosets. In: Box, H.O. (Ed.) *Primate Responses to Environmental Change.* London: Chapman and Hall.

Siegel, S. & Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioural Sciences.* Singapore: McGraw-Hill.

Smith, C., King, J.W.B. & Gilbert, N. (1962). Genetic parameters of British Large White pigs. *Anim. Prod.*, **4**, 128-143.

Shimada, M. K., Inoue-Murayama, M., Ueda, Y., Maejima, M., Murayama, Y., Takenaka, O., Hayasaka, I. & Ito, S. (2004). *Biochemical and Biophysical Research communications*, **316**, 1186-1190.

Smuts, B. B. (1987). Gender, aggression and influence. In *Primate Societies.* Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W. & Struhsaker, T. T. (Eds). Chicago, University of Chicago Press.

Sokal, R. R. &Rohlf, F. J. (1994). *Biometry: The Principles and Practice of Statistics in Biological Research.* W H Freeman.

SPSS for Windows, Rel. 11.0.1. (2001). Chicago: SPSS Inc.

Stevenson, M.F. & Poole, T.B. (1976). An ethogram of the common marmoset (*Callithrix jacchus*): general behavioural repertoire. *Animal Behaviour,* **24**, 428-451.

Stevenson, M.F. (1977). The common marmoset (*Callithrix jacchus jacchus*) as a model for ethological research. *Laboratory Animal Science* **27,** 895-900.

Stevenson-Hinde, J. & Zunz, M. (1978). Subjective assessment of individual rhesus monkeys. *Primates*, **19**, 473-482.

Stevenson-Hinde, J., Stilwell-Barnes, R. & Zunz, M. (1980). Subjective assessment of rhesus monkeys over four successive years. *Primates* **21**, 66-82.

Strang, G.S. & Smith, C. (1979). A note on the heritability of traits in pigs. *Anim. Prod.*, **28**, 403-406.

Suomi, S.J. (1991). Uptight and laidback monkeys. Pp 27-56 In: Brauth, S.E., Hall, W.S. & Dooling, R.J. (eds.) *Plasticity in Development.* Cambridge, MA: MIT Press.

Sussman, R.W. (2002). *Primate Ecology and Social structure, Volume 2: New world Monkeys.* Needham Heights, MA: Pearson Custom Publishing.

Tardiff, S. D, Harrison, M. L. & Simek, M. A. (1993). Communal infant care in marmosets and tamarins: relation to energetics, ecology, and social organisation. In Rylands, A.B. (Ed) *Marmosets and Tamarins: Systematics, Behaviour, and Ecology.* New York, Oxford University Press.

Taylor, W.J., Brown, D.A., Davis, W.L. & Laudenslager, M.L. (1997). Novelty influences use of play structures by a group of socially housed bonnet macaques. *Lab. Primate Newsletter*, **36**, 4–6

Trivers, R.L., & Willard, D.E. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science,* **179**, 90-92.

Utami, S.S., Goossens, B., Bruford, M.W., de Ruiter, J.R. & van Hooff, J.A.R.A.M. (2002). Male bimaturism and reproductive success in Sumatran orang-utans. *Behavioral Ecology* **13**, 643-652.

van Rhijn, J.G. (1973). Behavioural dimorphism in male ruffs *Philogamus pugnax* L. *Behaviou,r* **47**,153–229.

van Rhijn, J.G. (1983). On the maintenance and origin of alternative strategies in the ruff *Philomachus pugnax. Ibis*, **125**, 482–498.

Visalberghi, E. Fragaszy, D.M. (1995). The behaviour of capuchin monkeys, *Cebus apella,* with novel food: The role of social conflict. *Animal Behaviour* **49**, 1089-1095.

Visalberghi, E., Janson, C. H. & Agostini, I. (2003). Response toward novel objects and novel food in *Cebus apella. International Journal of Primatology,* **24,** 653-675.

Visser, E. K., van Reenen, C. G., van der Werf, J. T. N., Schiolder, M. B. H., Knaap, J. H., Barneveld, A & Blokhuis, H. J. (2002) Heart rate variability during a novel object test and a handling test in young horses. *Physiol. Behav.* **76**, 289-296.

Vitale, A, Santamaria, F. & Queyras, A. (1997). Response to a novel object by socially-housed common marmosets. In Pryce, C., Scott, L. & Schnell, C. *Marmosets and Tamarins in Biological and Biomedical Research.* Salisbury: DSSD Imagery.

Voland, E. (1978). Social play behaviour of the common marmoset (*Callithrix jacchus* Erxl.,1777) in captivity. *Primates*, **18**, 883-901.

Wade, J. & Arnold, A.P. (2004). Sexual differentiation of the zebra finch song system. Ann. N Y Acad. Sci., **1016**, 540-59.

Walters, J.R. (1987). Transition to adulthood In: Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W. & Struhsaker, T. T. (eds) *Primate Societies* pp. 358–369. Chicago, University of Chicago Press.

Ward, A.J.W., Thomas, P., Hart, P.J.B. & Krause, J. (2004). Correlates of boldness in three-spined sticklebacks (*Gasterosteus aculeatus*). *Behavioural Ecology & Sociobiology,* **55**, 561-568.

Weiss, A., King, J.E. & Figueredo, A.J. (2000). The heritability of personality factors in chimpanzees (*Pan troglodytes*). *Behavior Genetics* **30**, 213-221.

Williamson, D.E., Coleman, K., Bacanu, S-A., Devlin, B. J., Rogers, J., Ryan, N. D. & Cameron, J. L. (2003). Heritability of fearful-anxious endophenotypes in infant rhesus macaques: a preliminary report. *Biol. Psychiatry,* **53**, 284-291.

Wilson, D.S., Clark, A.B., Coleman, K. & Dearstyne, T. (1994). Shyness and boldness in humans and other animals. *Trends in Ecology and Evolution,* **9,** 442-446.

Wilson, D.S. & Yoshimura, J. (1994). On the coexistence of specialists and generalists. *American Naturalist* **144**, 692-707.

Yamamoto, M. E. (1993). From dependence to sexual maturity: the behavioural ontogeny of Callitrichidae. In Rylands, A.B (ed). *Marmosets & Tamarins: Systematics, Behaviour, Ecology,* 235-256. Oxford university Press; Oxford.

Yamamoto, M.E., Domeniconi, C. & Box, H. (2004).Sex differences in common marmosets (*Callithrix jacchus*) in response to an unfamiliar food task. Primates, **45**, 249-254.

Zhou, Q-Y & Palmiter R.D. (1995). Dopamine-deficient mice are severely hypoactive, adipsic and aphagic. *Cell,* **83**, 1197-209.