

Automated tracking of behavioural synchrony in cooperating marmosets

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Vasudha Kulkarni



Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, India.

Date: 27 March 2024

Under the guidance of

Supervisor: **Prof. Dr. Judith M. Burkart,**

Institute of Evolutionary Anthropology, University of Zürich
Switzerland

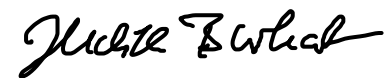
From June 2023 to March 2024

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Certificate

This is to certify that this dissertation entitled '**Automated tracking of behavioural synchrony in cooperating marmosets**' towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents work carried out by Vasudha Kulkarni at the Institute of Evolutionary Anthropology, University of Zürich, Switzerland under the supervision of Prof. Judith M. Burkart, Professor, Institute of Evolutionary Anthropology, University of Zürich, during the academic year 2023-2024.

Prof. Judith M. Burkart



Thesis Advisory Committee:

Prof. Judith M. Burkart

Professor, Institute of Evolutionary Anthropology
University of Zürich

Dr. Raghav Rajan

Assistant Professor, Department of Biology
Indian Institute of Science Education and Research, Pune

This thesis is dedicated to all my teachers and mentors.

Declaration

I hereby declare that the matter embodied in the report entitled '**Automated tracking of behavioural synchrony in cooperating marmosets**' are the results of the work carried out by me at the Institute of Evolutionary Anthropology, University of Zürich, Switzerland, under the supervision of Prof. Judith M. Burkart, and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.



Vasudha Kulkarni

Reg. no: 20191057

Date: 27th March 2024

Table of Contents

	Abstract	8
	Acknowledgements	9
	Contributions	11
Chapter 1	Introduction	12
	Synchrony: Basis of affiliative bonds	12
	Common marmosets as a model for studying human cognitive evolution	13
	Behavioural synchrony in non-human primates	15
	Mutual gaze in marmosets	16
	Challenge of automated pose estimation	17
	Quantifying behavioural synchrony	19
	Experimental tasks	21
	Aims and predictions of the project	21
Chapter 2	Methods	24
	Subject and housing	24
	Experimental arena	24
	Video and audio recording	25
	Experimental task	25
	Behavioural coding	28
	Video processing	28
	DeepLabCut projects	29
	Pose3D: Camera calibration	29
	Processing DLC detections	32
Chapter 3	Results	37
	Prosocial task results	37
	DeepLabCut model comparison and evaluation	38
	3D Camera Calibration Parameters	40
	DLC detections summary	40
	Gaze analysis during task	41
	Task kinematics during the individual task	44
Chapter 4	Discussion	46
	Bibliography	49
	Appendix	53

List of Tables

Tables	Legend	Page number
Table 1	Edge lengths of 3D reconstructed cuboids for 3D alignment	33
Table 2	Serial rotation angles along corresponding axes for 3D transformation	33
Table 3	Relative scaling factor between left and right 3D reconstructed space	33
Table 4	Statistical test results of dynamic time warping distance comparison	45

List of Figures

Figures	Legend	Page number
1	Social interactions in common marmosets	14
2	Automated pose tracking software	18
3	Illustration of parameter estimation in camera calibration	19
4	Illustration of DTW of 3D trajectories of a walking sequence	21
5	Experimental arena	24
6	Orientation of Raspberry Pi cameras in the experiment room	25
7	Prosocial and Individual sliding boards	26
8	Illustration of the No Partner control task	27
9	Timeline and order of experimental sessions	27
10	Body part labels on a marmoset for DLC training data	29
11	Camera calibration process in MATLAB	30
12	3D reconstruction and transformation of cuboids for 3D alignment	32
13	Illustration of calculation of angle between head vectors	34
14	Illustration of gaze intersection with sliding board	35
15	Rate of prosocial pulls across prosocial sessions	37
16	Rate of prosocial pulls in no partner control sessions	37
17	Number of pulls in prosocial consolidation sessions	38
18	Comparing the performance of DLC models	38
19	Accuracy of DLC labels	39
20	DLC model evaluation summary	39
21	Distribution of percentage of missing detections	40
22	Distribution of maximum number of continuous null detections	40
23	Distribution of median likelihood values across all body parts	41
24	Distribution of angle between head vectors	41
25	Elevation angle between head vector and gaze vector	42
26	Fraction of time the individual's gaze intersects the sliding board	43
27	Gaze intersection during prosocial sessions differentiated by role of the individuals	43
28	Measure of handedness and duration of task execution from hand detections	44
29	Dynamic time warping distance measure of task execution	45

Abstract

Interacting humans tend to align with each other at physiological, neural and behavioural levels. The degree to which behaviours in an interaction are patterned or synchronised in both timing and form is termed behavioural synchrony. The degree of behavioural synchrony often correlates with cooperation, prosocial behaviour, and social cognition in human interactions. Intense cooperation and proactive prosociality have convergently evolved in humans and cooperatively breeding common marmosets, which makes them a great model to study the mechanisms underlying social cognition. Marmosets regularly engage in cooperative tasks, but it is not known if they too exhibit behavioural synchrony to facilitate coordination. It is also challenging to define and objectively evaluate behavioural synchrony and posture imitation in non-human animals. Here, we conducted an experiment to study behavioural synchrony in marmoset dyads before and after they engage in a prosocial task to investigate the effect of prosociality on behavioural synchrony. We established a pipeline to extract trajectories of multiple marmosets from video recordings. As a part of the pipeline, we first tracked the body parts of marmoset dyads from different angles using DeepLabCut, a markerless, automated, machine-learning-based pose estimation tool. We then used MATLAB's stereo-camera calibration to reconstruct 3D coordinates of the marmoset body parts. With this pipeline, we examined the gaze direction and kinematics of marmoset hand movements during the task. We found that marmoset dyads mostly look in the same direction and exhibit distinct gaze patterns across different task conditions. We used dynamic time warping to analyse similarity in hand movements and found that individuals have more consistent patterns of task execution within themselves when compared to dyads. However, we did not find greater similarity in task kinematics of real dyads as compared to pseudo-dyads. Further studying the processes of synchronisation in freely moving marmosets will help us understand the overlap of proximate mechanisms regulating cooperation and social cognition in humans and marmosets.

Acknowledgements

It has been an honour and a privilege to work on this project with incredible scientists and marmosets for my master's thesis that is presented before you. This work wouldn't have been possible without Prof. Judith Burkart, for whom I am grateful for this incredible opportunity, her guidance and invaluable support throughout the year, and for always being as excited about the project as I was. I am very grateful to Nikhil Phaniraj for sharing his expertise and elevating this project, his comments on the thesis, and his patience and encouraging feedback as I navigated a steep learning curve. I would like to express my gratitude to all the Evolutionary Cognition Group members for the engaging discussions and support during my project – especially the caretakers, Hidir Sengül and Dominique Ziegler for their assistance during the experiments, Dr. Rahel Brügger and Konatsu Ono for advice on working with the marmosets, and Dr. Paola Cerrito and Adele Tuozzi for being the most supportive officemates they are. I would also like to thank Dr. Raghav Rajan for his comments; and Divyansh Gupta and Amogh Rakesh for the discussions that helped me implement some code much faster than I could have by myself. The funding I received from the A.H. Schultz Foundation allowed me to carry out this work in Zürich, for which I'm deeply grateful.

My academic journey has been largely shaped by teachers and mentors who shared their fascination for the subject and also fostered mine. My interest in animal behaviour and evolution was sparked by courses taught by Prof. Sutirth Dey and Dr. Anand Krishnan at IISER Pune, who made the study of the patterns in animal behaviour sound like a great adventure. Once I was determined to explore the subject, Prof. Raghavendra Gadagkar at IISc generously took me on, which enabled me to learn about the fundamentals by reading books, writing about them and weekly discussions with a luminary in the field. I learnt not only about animal cognition, evolutionary processes and brood parasitism, but this project was pivotal in developing my interests in social behaviour and led to an incredible internship the following summer with Prof. Sylvia Cremer at ISTA, where I discovered my passion for experimental work. Through subsequent research projects with Dr. Raghav Rajan, I learnt important methods in data analysis and programming, which solidified my interests in combining quantitative methods and empirical behavioural studies to explore social interactions in animals. I am profoundly grateful to all my mentors for the sustained interactions while developing the projects and their detailed feedback on my work, which has brought me to this level.

I would also like to thank Dr Deepak Barua, Dr Nishikant Subhedar, Dr Pooja Sancheti, Dr. Chaitra Redkar and Dr Shalini Sharma for their remarkable courses through which I developed an in depth understanding of the field and an overall appreciation for interdisciplinary studies. I'm very grateful to have studied in IISER Pune, which created an atmosphere of learning and discovery, and for the KVPY fellowship, which supported me throughout my undergraduate studies.

The last five years have also been incredibly rewarding in terms of the people who have come into my life and made it all the better. I extend my gratitude Mariann Strauli for sharing her home with me, for the better part of a year, when I desperately needed one. I'm filled with appreciation for the

student community at IISER Pune, I learnt more from my peers than I did on my own. I would like to extend my sincere gratitude everyone who has supported me and kept me together over the memorable last five years – Aditya, both Amoghs, Akash, Krishna, Varun, Neev, Aniketh, Vatsal, Garvit, Jezer, Pallav, Sanjana, Ritvee, Likhith, Shree Hari, Kaustav, Saismit, and Divyansh, in conversations with whom I grew as a person and as a scientist.

Lastly, I wouldn't be where I am today without the support and encouragement of my family. I'm very grateful to my sister, Aparna, for always being in my corner, and my parents, Appaji and Amma, for always encouraging my questions and trusting me to make my own decisions.

Contributions

Contributor name	Contributor role
Judith Burkart, Nikhil Phaniraj, Vasudha Kulkarni	Conceptualization Ideas
Judith Burkart, Nikhil Phaniraj, Vasudha Kulkarni	Methodology
Markus Marks	Software
Judith Burkart, Nikhil Phaniraj, Vasudha Kulkarni	Validation
Vasudha Kulkarni	Formal analysis
Vasudha Kulkarni	Investigation
Judith Burkart	Resources
Vasudha Kulkarni	Data Curation
Vasudha Kulkarni	Writing – original draft preparation
Judith Burkart, Nikhil Phaniraj, Vasudha Kulkarni	Writing – review and editing
Vasudha Kulkarni	Visualization
Judith Burkart, Nikhil Phaniraj	Supervision
Judith Burkart	Project administration
Judith Burkart	Funding acquisition

Chapter 1: Introduction

Synchrony: Basis of affiliative bonds

When two people interact, their interaction develops a temporal structure as they tune in and synchronise with each other at physiological, neural and behavioural levels. The bio-behavioural synchrony that develops between two people is hypothesised to be the fundamental organising principle in establishing and maintaining selective and enduring affiliative bonds between humans (Feldman, 2012). Several studies on interacting mother-infant, therapist-patient and romantic couple dyads have shown that they synchronise their heart rate, breathing rate and arousal level, all of which are regulated by the autonomic nervous system (Stratford *et al.*, 2012; Ferrer and Helm, 2013; Creaven *et al.*, 2014; Palumbo *et al.*, 2017). Conversing partners over time also morph their pronunciation, vocal range, syntactic structure and lexical choice to converge with that of the other individual, in a phenomenon known as social vocal accommodation (Brennan and Clark, 1996; Babel, 2012; Healey *et al.*, 2014; Ruch *et al.*, 2018). Interacting partners also tend to couple the timing and content of their neural activity, which helps in successful communication (Stephens *et al.*, 2010; Gvirts and Perlmutter, 2020). Following gaze direction is also a vital component of social interactions and is important for establishing joint attention (Emery, 2000; Shepherd, 2010). In addition, these dyads also tend to imitate each other's postures and movements, following one another in their hand gestures or the position of their feet, thus also exhibiting behavioural synchrony (Chartrand and Bargh, 1999; Shockley *et al.*, 2003).

Behavioural synchrony can be defined as the degree to which the behaviours in an interaction are non-random, patterned, or synchronised in both timing and form (Bernieri and Rosenthal, 1991). Delaherche *et al.* expand the definition of behavioural synchrony as the *dynamic* and *reciprocal* adaptation of the temporal structure of behaviours between interactive partners (Delaherche *et al.*, 2012). Bernieri *et al.* 1988 (Bernieri, 1988) defined three types of behavioural synchrony: interaction rhythms, which involve a sequence of identical behaviours between partners over time (e.g., levels of engagement in mother-infant interaction); simultaneous behaviours, where interacting partners exhibit identical behaviour (e.g., pose imitation and mimicry); and behavioural meshing, wherein interacting partners behave in a way that is complementary and forms a meaningful whole. Behavioural synchrony can also be characterised by temporal synchrony, where behaviours are synchronised in time; local synchrony, where individuals occupy the same space simultaneously; and allelomimicry, which involves exhibiting similar behaviours simultaneously (Duranton and Gaunet, 2016). The intentions of the actors can be used to categorise behavioural synchrony into intentional and incidental synchrony (Rennung and Göritz, 2016). Intentional synchrony is when individuals coordinate to achieve a common external goal. Incidental synchrony is the nonconscious mimicry of body postures or facial expressions, which is also known as the Chameleon effect. This imitation is hypothesised to occur spontaneously through a perception-behaviour link; that is, the mere perception of an individual's action increases the probability of that action in the observer (Chartrand and Bargh, 1999).

Several studies have shown that individuals aligning with each other through synchronised movements are better at cooperation, prosociality and coordinating with each other to achieve common goals. In a study by Hove and Risen, participants were made to tap their fingers synchronously or asynchronously with the other participants (Hove and Risen, 2009). They reported that individuals who tapped their fingers synchronously reported a higher affiliation score than the asynchronous dyads. Tunçgenç *et al.* similarly showed that synchronous movement between groups increased outgroup bonding (Tunçgenç and Cohen 2016). In an experiment by Macrae *et al.*, participants were made to place their forearms on the table and raise their hands

up and down with the experimenter's hand in a synchronous or asynchronous condition while the experimenter recited some common words (Macrae *et al.*, 2008). Participants who were synchronised with the experimenter were better at facial recognition of the experimenter and were able to recall a greater number of words than the unsynchronised participants. Similarly, Valdesolo *et al.*, showed that synchronised individuals cooperated better in a joint action task due to their increased perception sensitivity (Valdesolo *et al.*, 2010). Finally, Reddish *et al.*, conducted a study where participants were made to dance synchronously or asynchronously in groups of three with a beat they could hear through headphones (Reddish *et al.*, 2013). Participants who danced synchronously were more prosocial and contributed more to an economic public goods game, in contrast to those who danced asynchronously. These studies highlight the important role of behavioural synchrony in promoting cooperation and social cohesion.

There are several theories on how behavioural synchrony enhances cooperation, prosociality, and coordination within a group (Mogan *et al.*, 2017). One theory suggests that behavioural synchrony and mimicry are ways of signalling similarity that blur the self-other boundaries, allowing for increased empathy towards and cooperation with another individual (Hove, 2008). Another theory posits that behavioural synchrony is a way of signalling a willingness to cooperate within a group, which increases the proclivity of group members to cooperate. In addition, behavioural synchrony enhances social attention, which facilitates improved coordination during joint tasks (Macrae *et al.*, 2008). Lastly, Wheatley *et al.* suggest that performing the same actions at the same time aligns the neural activities and representation of individuals, which allows them to 'get on the same page', thus improving cooperation (Wheatley *et al.*, 2012). Lang *et al.* conducted an experiment to test some of these theories and showed that self-other overlap and perceived cooperation mediated the effects of synchrony on interpersonal affiliation, whereas beta-endorphin release mediated the effects on cooperation between participants (Lang *et al.*, 2017). Thus, various mechanisms could be mediating the effects of behavioural synchrony on cooperation and social bonding.

Despite several studies demonstrating a positive correlation between biobehavioral synchrony and stronger affiliative bonds, the direction of causation remains a subject of debate. One perspective suggests that individuals with strong affiliations may share similar representations of the world, leading to synchronised actions. Conversely, it has also been suggested that synchronous actions themselves foster deeper affiliative bonds. Additionally, while synchrony is theorised to enhance empathy, Chartrand and Bargh's work indicates that highly empathetic individuals exhibit the Chameleon Effect more prominently (Chartrand and Bargh, 1999). Thus, synchrony and affiliation are intricately intertwined, with the relationship between them warranting further exploration.

Common Marmosets as a model to study the evolution of human cognition

Humans are unique in their higher-order cognitive abilities, which allows us to recognise shared intentions and thus coordinate across large spatial and temporal scales to achieve common goals and build on our cumulative knowledge (Tomasello *et al.*, 2005). In order to understand the primate origins of human nature, we can first look towards great apes because we share several similarities with them that can be explained by shared ancestry. But we are substantially different from them in certain features such as longer juvenile periods, larger, energy-intensive brains, and higher cognitive abilities (Antón *et al.*, 2014). Cognitive tests between chimpanzees, orangutans and 2.5-year-old human children revealed that human children performed similarly in spatial cognitive tests, such as spatial memory, discriminating quantities and grasping causality. However, the children were better at socio-cognitive tasks – such as social learning, communication and

theory of mind – as compared to adult chimpanzees and orangutans (Herrmann *et al.*, 2010). Thus, the intrinsic gap between the human mind and the ape mind is social.

The Cooperative Breeding Hypothesis of human evolution posits that the evolution of allomaternal care provided the prosocial motivation and high social tolerance to acquire species-specific cognitive abilities, such as effective social learning, shared intentionality and cumulative culture (Burkart *et al.*, 2009; Burkart and Finkenwirth, 2015). In order to understand the evolutionary mechanisms of human social cognition, we can study the cognitive effects of cooperative breeding in a primate species where it has evolved convergently, such as the Common marmosets (*Callithrix jacchus*) and contrast it with the cognitive abilities of an independently breeding sister taxon, such as squirrel monkeys (*Saimiri sciureus*).

Common marmosets are a neotropical primate species belonging to the Callitrichidae family. They are cooperative breeders, wherein individuals who are not the parents help care for the infants. Like humans, marmosets exhibit proactive prosociality and group-level coordination. They live in groups of 6-15 individuals, usually composed of only one socially dominant breeding pair. Other individuals in the group help care for the infants by carrying them, provisioning them, exhibiting anti-predator vigilance, and territorial defence (Burkart *et al.*, 2009; Guerreiro Martins *et al.*, 2019). They are highly cooperative within groups and regularly groom each other, share information about the location of food and feed in proximity to each other, as shown in Figure 1 (Finkenwirth *et al.*, 2016). Moreover, they breed easily in captivity, become sexually mature at 1.5 years and produce twins or triplets twice a year. Their small body size facilitates easy handling and maintenance. Thus, marmosets are an ideal primate model for studying the evolution of human cognition because of the ease of maintenance and their evolutionary relationship with humans (Burkart and Finkenwirth, 2015).

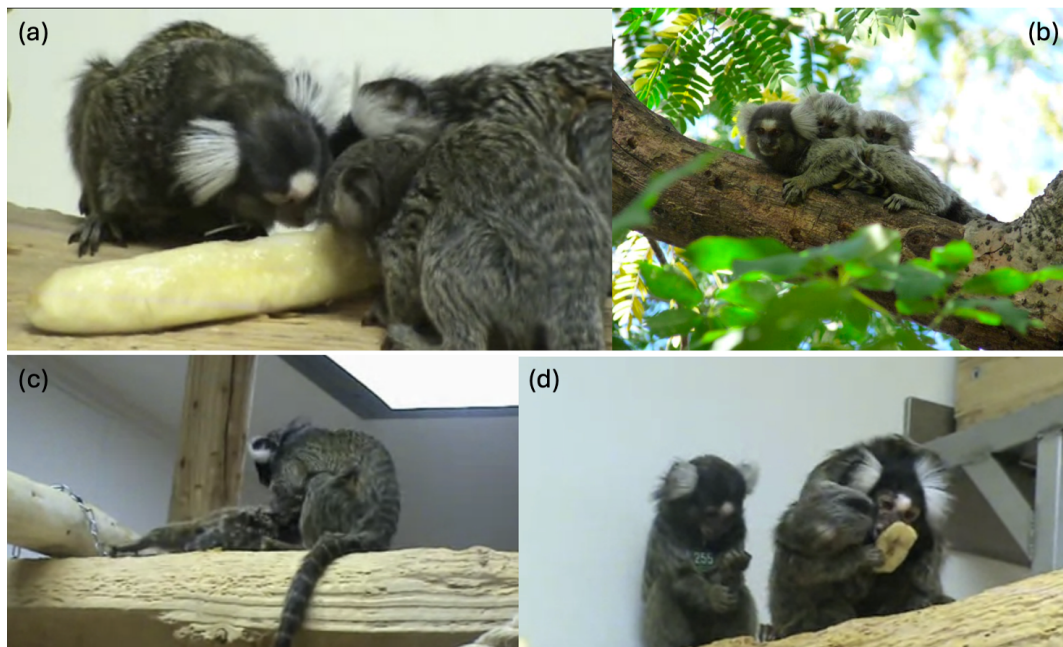


Figure 1. Social interactions in common marmosets. a) Social tolerance b) Infant carrying c) Grooming each other d) Food sharing in common marmosets [Images from Common Marmoset Care (<https://www.marmosetcare.stir.ac.uk/>)]

Marmosets regularly engage in refined cooperation and joint action with other individuals in the group (Snowdon, 2001). They also participate in cooperative problem-solving, and when performing a cooperative task, they represent not only their own task and actions but also their partner's, i.e. they show action co-representation in the Joint Simon task (Miss and Burkart, 2018). Marmosets also have a vast repertoire of vocal communication for a nonhuman primate, and they show similar communicative features as humans, such as a babbling phase in immatures (Snowdon and Elowson, 2001), cooperative vocal turn-taking (Takahashi *et al.*, 2013), and vocal dialects in groups (Zürcher and Burkart, 2017). This also makes them a great system for studying language evolution.

As cooperative breeders, marmosets in the group coordinate and cooperate closely to care for the infants. They show a group-level similarity in personality (Koski and Burkart, 2015) and tend to be in tune with the group at communicative, hormonal and behavioural levels. Closely bonded individuals show synchronised fluctuation of oxytocin, and breeding pairs with strong oxytocin synchronisation tend to care better for their infants (Finkenwirth and Burkart, 2017, 2018). Marmosets also show contagious behaviour with arousal states (de Oliveira Terceiro *et al.*, 2021)(de Oliveira Terceiro *et al.*, 2021b)(de Oliveira Terceiro *et al.*, 2021) and gnawing and scent marking, which could promote group coordination by facilitating activity transitions and state matching (Massen *et al.*, 2016).

In marmoset dyads, individuals also take turns being vigilant while the other one is feeding, and they calibrate their vigilance to their partner's risk level (Brügger *et al.*, 2022). This turn-taking in vigilance behaviour can be modelled as anti-phase synchrony (Phaniraj *et al.*, 2023). These observations in marmosets, along with their high social tolerance and social monitoring, indicate that synchronisation and tuning-in could be the underlying mechanisms for cooperation that differentiate the social abilities of marmosets from independently breeding sister taxa. Studying the processes of synchronisation in marmosets will help us understand the overlap of proximate mechanisms regulating cooperation and social cognition in humans and marmosets.

Behavioural synchrony in non-human primates

Previous studies that have looked at behavioural synchrony in non-human primates have considered a broad, higher-order behavioural categorisation such as foraging, travelling, sleeping, grooming and so on, and have used scan sampling and focal sampling to quantify the number of individuals engaged in a particular activity. Behavioural synchrony in Chacma baboon groups increased with the increase in the number of pregnant females (King and Cowlshaw, 2009). Daoudi-Simison *et al.* studied behavioural synchrony between mixed-species groups of squirrel monkeys and capuchins and found that intra-specific behavioural synchrony was higher than random aggregates but did not find any inter-specific synchronisation (Daoudi-Simison *et al.*, 2023). Nishikawa *et al.* observed that activity synchrony in Japanese macaque females was mediated by spatial cohesion and type of activity (Nishikawa *et al.*, 2021). Such activity synchronisation in group living species increases the effectiveness of anti-predator defence due to the dilution effect (which decreases an individual's chance of being caught by a predator) and the increased effectiveness of vigilance against predators (Duranton and Gaunet, 2016).

A couple of studies have examined the kinematics of action execution in non-human primates, using frame-by-frame manual identification of body parts. One study looked at the patterns of motor mimicking in tool-based nut-cracking action in chimpanzees (Fuhrmann *et al.*, 2014). They analysed several videos of a model

animal demonstrating the nut cracking and an observer chimp mimicking the action and demonstrated a synchronised motor pattern that was transmitted unidirectionally from the model to the observer. Similarly, Voelkl and Huber trained a model marmoset to open a canister using its mouth and trained several observer and non-observer individuals to open the same canister (Voelkl and Huber, 2007). They quantified the movement pattern of manually labelled face and hand markers and found that observer individuals copied the same action in contrast to the non-observer individuals who independently found the same solution to the task (i.e. opening the canister with their mouth).

The above mentioned studies have looked at macro-scale activity categories or inspected action mimicry at a fine scale in a particular problem-solving context but with a very labour-intensive method. To my knowledge, there has been no experiment that has studied the phenomenon of pose imitation or perception-behaviour link in non-human primates. In this context, our goal was to study behavioural synchrony in cooperating marmosets at a fine temporal resolution using automated pose estimation software, which extracts spatial data in three dimensions.

Mutual gaze in marmosets

Humans are incredible at inferring emotions and intentions from another individual's gaze. Gaze following is an important component of social interactions – it allows us to focus on the object of attention and coordinate with another individual to achieve a task (Emery, 2000). In most other non-human primates, especially macaques, looking directly into an individual's eyes is perceived as a threat, and they have a strong gaze aversion response (Perrett and Mistlin, 1990). Unlike the old-world primates, marmosets lack gaze aversion and use mutual gazing and gaze following to facilitate social learning and joint action tasks. Previous studies have shown that marmosets can geometrically infer the position of reward, among nine choices, from a human's gaze standing 1 metre away (Burkart and Heschl, 2006). They also engage in mutual gazing while coordinating a joint action task (Miss and Burkart, 2018) and use it to establish joint attention (Spadacenta *et al.*, 2019). This mutual gazing, along with prosociality, turn-taking and tuning in, is hypothesised to be a part of the marmoset interactional engine, which facilitates flexible communication in marmosets (Burkart *et al.*, 2022).

Marmosets have a very narrow oculomotor range of 10° (Mitchell *et al.*, 2014) as compared to macaques (40° - 50°) and humans (55°). Given the small size of their head (~ 3.5 cm) and the forward-positioned eyes, marmosets use head movements to orient their gaze rather than rapid eye movements. They can turn their heads towards an auditory stimulus at a maximal speed of $1000^\circ/\text{s}$ for an amplitude greater than 120° , whereas humans could turn their heads at a maximal speed of $600^\circ/\text{s}$ for a rotation of 120° (Guitton and Volle, 1987; Pandey *et al.*, 2020). Thus, the direction of gaze in marmosets can be inferred from the direction of their head with a 10° error margin. A recent study by Xing *et al.* used automated 3D detection of head features of marmosets to show interesting gaze patterns in familiar and unfamiliar dyads (Xing *et al.*, 2024). Unfamiliar dyads showed increased levels of gaze monitoring and a higher probability of recurrent gaze towards the unfamiliar partner, while the familiar dyads engaged more in joint gazing when they were near and had a higher probability of taking turns looking at each other. Building on these studies, it would be interesting to explore how marmosets use gaze following while coordinating in a cooperative task.

The challenge of automated pose estimation

Analysing behavioural data using video recordings is a crucial part of studying behavioural synchrony. High-speed video recordings of behavioural experiments can yield a wealth of fine-grained data about the behaviour and lead to new insights, such as the discovery of tap dancing in songbirds (Ota *et al.*, 2015). Historically, the collected data was annotated and analysed manually, which is laborious, time-consuming, and fallible. Estimating the pose of an animal is the first step towards classifying behaviour from the temporal patterns of postures in videos. However, given the rapid data acquisition rate in recent years, the prevalent method to estimate postures was to use distinct, reflective markers to highlight the area of interest and use marker-based tracking software to track body parts accurately (Vargas-Irwin *et al.*, 2010). However, using markers can be distracting for the subjects. Markerless video tracking software overcomes this limitation and is capable of accurate and automatic pose estimation by using deep neural networks (Moro *et al.*, 2022).

A significant challenge in studying pose synchrony and imitation in non-human primates is a lack of an objective, consensus method to identify and classify them. When reviewing the various methods of quantifying behavioural synchrony in humans, a prominent early approach for human coding of synchrony is the pseudo-synchrony paradigm used in Bernieri 1988 (Bernieri, 1988), where synchrony was manually coded by untrained raters using gestalt judgments for a) genuine interactions between mother-infant dyad and b) pseudo-interactions created using video clips of the genuine pair re-combined in random order. The study used pseudo-synchrony rated by multiple coders as the baseline and compared the ratings of genuine interactions with this to show that genuine interactions of mothers-infant dyads had higher synchrony levels as compared to interaction with an unfamiliar child.

More recent studies on interpersonal synchrony in humans have used video-based tracking systems like Motion Energy Analysis (MEA) (Ramseyer and Tschacher, 2010) or OpenPose (Cao *et al.*, 2017), shown in Figure 2a. MEA is a pixel-differencing technique that automatically calculates the change in grayscale pixels between consecutive video frames, which are encoded as movements of individuals when the video recorder is fixed and the background and lighting remain unchanged. This technique is robust to different kinds of tasks, but it is challenging to track multiple individuals or in unstable light or background conditions. OpenPose is a markerless pose estimation method for humans which uses computer vision and deep learning to detect the coordinates of joint parts automatically. It can track multiple individuals and has reduced sensitivity to noise as compared to MEA, but it only recognises 2D coordinates (Fujiwara & Yokomitsu, 2021). In a similar study on synchronised body motion between conversing participants, the experimenters used Microsoft Kinect software to extract 3D coordinates of body joints (Gaziv *et al.*, 2017). Importantly, none of these software generalise well to non-human species.

With advancements in computer vision, machine learning, image recording and processing software, manual behavioural coding and classification have been augmented by automated markerless tools (Luxem *et al.*, 2023). Several open-source machine-learning-based pose-estimation tools for multi-animal videos have been developed in the past few years, such as DeepLabCut (Mathis *et al.*, 2018), SLEAP (Pereira *et al.*, 2022), DeepPoseKit (Graving *et al.*, 2019), and Trex (Walter and Couzin, 2021) to name a few. There have also been tools that extract poses or behaviours in 3D, such as BKIND-3D (Sun *et al.*, 2023), LiftPose3D (Gosztolai and Ramdya, 2022) and 3D-UPPER (Ebrahimi *et al.*, 2023), but are limited by required volume and type of training dataset, single animal videos or lack of generalisation to non-model organisms.

DeepLabCut (DLC) is an open-source, markerless pose estimation algorithm based on transfer learning (it uses a pre-trained neural network and trains it with a small, supervised dataset) and convolutional neural networks that detect the geometric configuration of body parts, thus automating tracking and pose estimation of animals in laboratory conditions as shown in Figure 2b (Mathis *et al.*, 2018). DLC uses a bottom-up approach, where it first localizes key points, groups detections and then assigns individual IDs. This approach allows us to study the whole-body posture of individuals. A small number of manually annotated frames (~ 200) is sufficient to train the neural network to track user-defined features of individuals in any recording. DLC was extended to Multi-animal DLC (maDLC) to identify and track multiple individuals in a video (Lauer *et al.*, 2022). We also chose to work with DLC because the maDLC networks have been pre-trained on a large marmoset video database, and DLC stands out for its good community support from the developers and users.

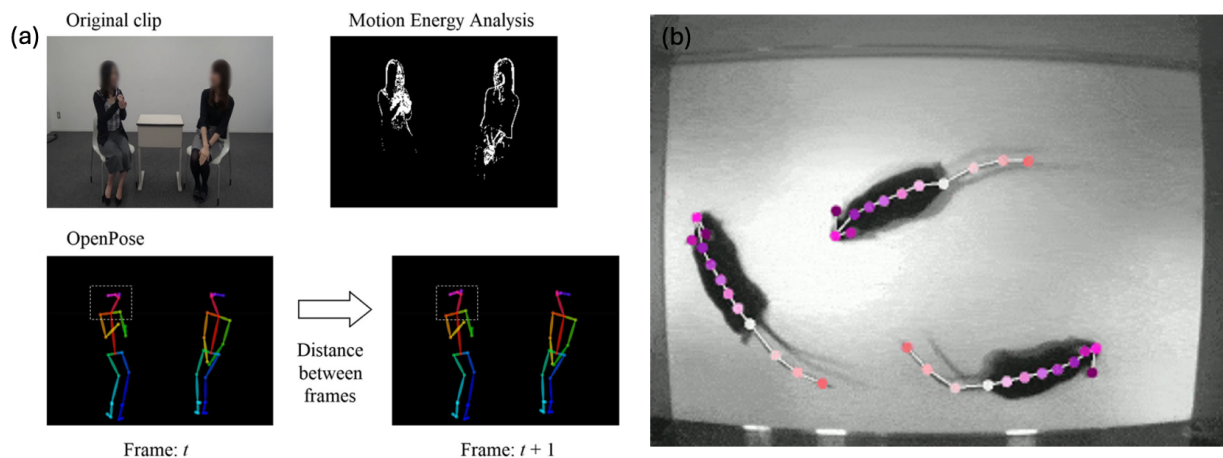


Figure 2. Automated pose tracking software. a) Examples of MEA and OpenPose (Fujiwara & Yokomitsu, 2021) b) Multi-animal pose detection by DeepLabCut (<https://deeplabcut.github.io/DeepLabCut/README.html>)

Multi-animal DLC can be used to extract the position of body parts over time to obtain time series data of the trajectories of the individual's body parts. Using this quantitatively characterised behavioural time series, we can extract relationships in the movement behaviour and body postures of the two individuals. This time series data could also be used to quantify synchrony and further correlate it with the extent of cooperation in the task and variation in the interdependence between the dyads.

Once the trajectories of the features have been extracted, in order to get a 3D reconstruction of these movements, we need to calibrate the cameras to estimate certain parameters that allow us to take the 2D coordinates to 3D coordinates. 3D reconstruction usually requires multiple cameras (at least two cameras) positioned and synchronised in a particular manner. The cameras must be positioned in a stereo manner with a similar field of view. The camera calibration process involves estimating intrinsic parameters (focal length and tangential and radial distortions of the camera lens) and extrinsic parameters (relative rotation and translation of cameras with respect to each other) using an algorithm to detect checkerboard corners of known dimensions. These parameters are then applied to the 2D coordinates from multiple cameras to first undistort and then triangulate them to get a single 3D coordinate.

There are several 3D camera calibration tools, such as DeepLabCut3D (Nath *et al.*, 2019), Anipose (Karashchuk *et al.*, 2021), Argus (Jackson *et al.*, 2016), EasyWand (Bluhm and Hedrick), DLTDv (Hedrick,

2008) and MATLAB StereoCameraCalibrator (Bouget, 2022). The first three programs use Python’s open-source computer vision library, OpenCV, which performed poorly in recognising checkerboard corners from our calibration trials. EasyWand and DLTDv use a different method, which involves annotating the coordinates of an object in 3D space. Fortunately, MATLAB’s StereoCameraCalibrator, which makes use of the MATLAB Computer Vision toolbox, is easy to use and performs well with corner detection. It requires 10-50 pairs of images of the checkerboard (at different distances and angles from the cameras) as input to calculate the extrinsic and intrinsic parameters of the cameras (Figure 3). These parameters are then used to undistort and transform the 2D coordinates of body parts from 2D maDLC projects and reconstruct them in 3D space. The efficacy of the calibration is determined by the reprojection error measured in pixels. Pose3D (Sheshadri *et al.*, 2020) is a semi-supervised pipeline that facilitates camera calibration in MATLAB and 3D reconstruction from multiple DLC detections.

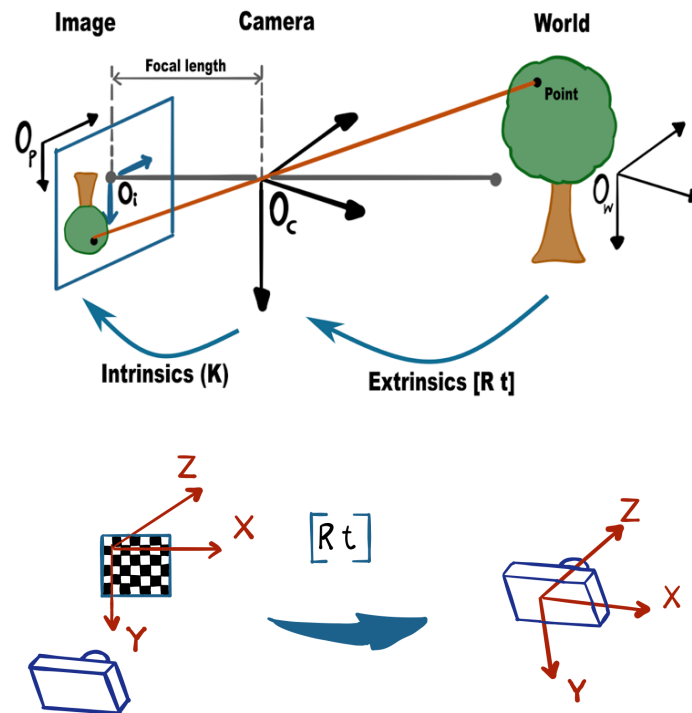


Figure 3. Illustration of parameter estimation in camera calibration. Extrinsic parameters define the position and angle of the camera in the real world, and intrinsic parameters contain information about the focal length, distortion and skew arising from the camera lens. These parameters are estimated using corners detected from a checkerboard of known dimensions [Adapted from MATLAB documentation]

Quantifying behavioural synchrony

Synchrony can be described as the dynamic and reciprocal adaptation of the temporal structure of behaviours between interacting individuals (Delaherche *et al.*, 2012). Once we have extracted movement trajectories of marmosets, the analysis of interpersonal synchrony involves obtaining time series data with a fixed time interval and performing a time series analysis to test for synchrony. The time series of the position and movement of two individuals can be analysed to estimate the extent of synchrony. A simple way of

quantifying synchrony is to estimate the cross-correlation, which is an extension of Pearson's correlation to time series data (Tschacher *et al.*, 2014).

Studies analysing behavioural synchrony in humans using experimental setups often tend to have individuals perform rhythmic activity, such as tapping their fingers, swinging a pendulum or rocking their chair, to create a synchronised condition (Schmidt and Turvey, 1994; Richardson, 2005; Hove and Risen, 2009). Schmidt *et al.* conducted an experiment to study postural synchrony between participants who told each other knock-knock jokes, and the cross-spectral analyses brought out the rhythmic nature of their interactions (Schmidt *et al.*, 2012). Such time series data is one-dimensional and, fulfils the assumptions of continuity periodicity and produces a stationary and regular time series, such that spectral methods of time series analysis like Fourier spectral analysis, cross-correlation and coherence can be used to evaluate synchrony (Shockley *et al.*, 2003). Different time series analysis methods are not perfectly convergent; rather, they quantify different aspects of synchrony, such as the strength of synchronisation during total or intervals of interaction or the frequency of synchronisation (Schoenherr *et al.*, 2019).

Several studies on physiological synchrony use Windowed Cross Correlation (WCC) to quantify synchrony, which analyses correlations between variables without the assumption that they are stationary, but it is still limited to one-dimensional data (Boker *et al.*, 2002). The advantage of WCC is that the direction of lag helps us identify the direction of synchrony, i.e., the leader and the follower in the interaction. Multi-dimensional, fine-grained postural data with irregular time series sensitive to lower dimension projection and distortion requires more sophisticated methods of analysis.

Recurrence analysis is a particularly well-suited method to quantify posture synchrony (Marwan *et al.*, 2007; Coco and Dale, 2014). It involves creating a plot of all the time points when the two variables had the same value or were in recurrent states. From this plot, we can quantify the percentage of time they spent in the same state, and the properties of diagonally aligned points are informative of periods of synchrony between two time series. The average length of the diagonal line in the plot tells us the average period of continuous attunement of the system, and the histogram of the diagonal line lengths can be used to obtain the entropy of the system (Delaherche *et al.*, 2012). Cross-recurrence quantification analysis (CRQA) has been used to show that dyads show greater postural synchrony when they're conversing with each other and that this interpersonal synchrony is mediated by convergent speaking patterns (Shockley *et al.*, 2003, 2007). Moreover, CRQA can be applied to multi-dimensional data, as shown by Wallot *et al.*, where they used multi-dimensional recurrence quantification analysis (MdrQA) to analyse interpersonal dyadic synchrony as well as global group-level synchrony while individuals were instructed to build multi-step origami constructions (Wallot *et al.*, 2016).

In order to study the similarity of task execution across individuals, we used dynamic time warping (DTW), which is a time series analysis algorithm that can be used to find an optimal alignment between two time series that may vary in speed (ed. M Müller, 2007). For instance, it can be used to analyse the similarity in gait of two individuals who are walking at different speeds (Figure 4). The algorithm can analyse multidimensional time series and gives a distance metric as an output, which is inversely related to the similarity between two time series.

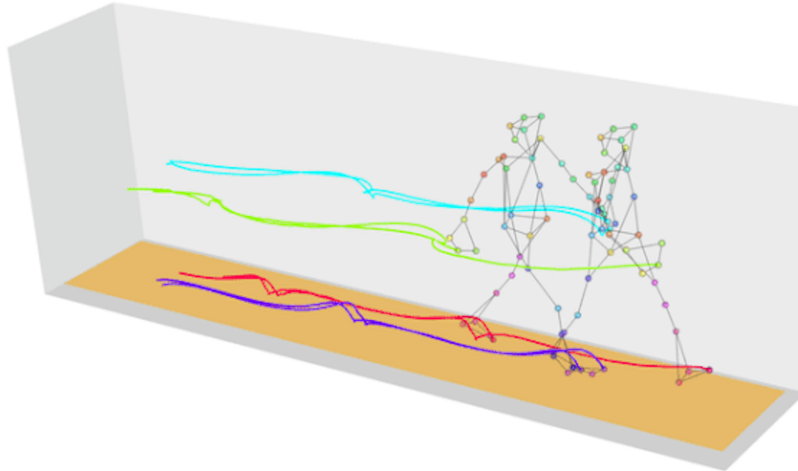


Figure 4. Illustration of dynamic time warping of 3D trajectories of a walking sequence. [Adapted from (Olsen *et al.*, 2018)]

Experimental tasks

In order to study behavioural synchrony in marmosets, we devised an experiment with different task conditions where marmosets would need to coordinate to different degrees and included a 2-minute period before and after the task where the marmosets are unrestricted and can interact freely. The **Prosocial task** consisted of a long sliding board with two food bowls on either end and a single handle on one end such that one individual could pull the board prosocially for the other individual to get the food treat. The handle was detachable such that it could alternate between the two individuals, and they could reciprocate. We could then quantify the baseline interpersonal behavioural synchrony of the dyad before the task and the difference in synchrony after the task and correlate it with their prosociality during the task.

As a control for the Prosocial task, we also devised an **Individual task** with separate sliding boards with food bowls such that the marmosets could pull the handle and get food rewards for themselves. In the Prosocial task, individuals had to coordinate with each other such that they were both in front of the task, with one individual ready to take the food when the other individual pulled the board. Meanwhile, in the individual task, they didn't have to focus on each other and only focused on the food reward for themselves. We also had a third task, the **Joint reward task**, with the long sliding board where only one individual could pull the board, but both individuals received the reward, unlike the Prosocial task where only the individual without the handle received the reward. In contrast to the Prosocial task, the second individual also received the food reward, but only as a consequence of the first individual pulling the board for itself. This task would create a higher positive affect between the dyad because they are still coordinating with each other to get food rewards as compared to the Individual or unsuccessful Prosocial tasks.

Aims and predictions of the project

The goal of my thesis was to set up an automated feature detection and 3D reconstruction pipeline for freely moving marmosets and use this framework to investigate gaze following, similarity in task kinematics and behavioural synchrony in marmoset dyads participating in a prosocial task. In order to set up the pipeline, we first implemented multi-animal DeepLabCut projects to extract coordinates of marmoset body parts from different angles. We then used camera calibration to obtain 3D coordinates from two sets of 2D coordinates

and transformed the 3D coordinates to align them to the real world. We used this framework to examine interactions between marmosets during the task.

Our goal was to study the gaze direction of the individuals and similarity in task kinematics across the prosocial, joint reward and individual tasks. The measures we computed and the trends we predicted for each component are as follows –

Gaze direction

1. Angle between head vectors
 - 1.1. Estimation: Calculate the angle between the vectors defined by mid-point between the ears to the nose.
 - 1.2. Prediction: Angle between head vectors indicates that both individuals looking in the same direction for most of the time, across different task conditions.
2. Fraction of time the gaze intersects with the task board
 - 2.1. Estimation: Calculate the fraction of time the gaze vector intersects their own task board versus the other individual's task board.
 - 2.2. Prediction: Individuals look at the other individuals' task board for longer, and at their own board for shorter time during the prosocial task as compared to joint reward and individual task.
3. Fraction of time the gaze intersects with the task board during prosocial session
 - 3.1. Estimation: Calculate the fraction of time the gaze vector intersects their own task board versus the other individual's task board and differentiate the status of the individual on whether they are providing or receiving the food reward during successful and unsuccessful prosocial trials.
 - 3.2. Prediction:
 - 3.2.1. If the marmosets are simply following the food reward, pullers would spend more time looking at the other individual's board as compared to the receivers, but if they are coordinating during the task, they would look at each other's board for similar amounts of time.
 - 3.2.2. Both individuals would spend more time engaging with the task board during the successful prosocial trials as compared to the unsuccessful ones.

Task kinematics

1. Preference for the use of one hand over the other
 - 1.1. Estimation: Calculate the log of the ratio of path lengths of left hand to the right hand during each trial as a measure of handedness.
 - 1.2. Prediction: Individuals would develop a preference for the use of one hand over the other across trials in a session or across sessions.
2. Efficiency of executing the task
 - 2.1. Estimation: Calculate the duration of the execution of each pull during the individual task.
 - 2.2. Prediction: Duration of the pull would decrease across trials in a session and across sessions as individuals became more efficient in executing the task.
3. Similarity in task kinematics
 - 3.1. Estimation: Perform dynamic time warping on 3D trajectories of hand detections during the task execution to measure the similarity of trajectories and compare the pulls of the same individual and pulls of two individuals in a dyad and compare the pulls of real dyads with those of pseudo dyads.
 - 3.2. Prediction:

- 3.2.1. The trajectories of the pulls of the same individual across trials would be more similar when compared to the pulls of two individuals in a trial.
- 3.2.2. The trajectories of the pulls of real dyads in a session would be more similar when compared to the pulls of pseudo dyads.

Chapter 2: Methods

Subjects and housing

For this experiment, we worked with nine individuals from four marmoset groups (two groups of mated pairs, one pair of siblings and three siblings from one group). Marmosets are housed in 1.8m x 2.4m x 3.6m heated indoor enclosures with access to a 1.8m x 2.4m x 4.6m outdoor enclosure. The enclosures have sufficient enrichments, and the floor is covered in bark mulch. They are provided with vitamin-enhanced rice flour mash in the morning, vegetables at noon, and snacks (cheese, nuts, cookies, etc.) in the afternoon every day. Water and food pellets are available ad libitum. The lighting is maintained at a 12-hour day-night cycle. All experiments were carried out in accordance with Swiss legislation and licensed by the Kantonales Veterinaramt Zürich (licence number: ZH223/16; degree of severity: 0).

Experimental arena

The experiments were conducted in the experimental room (Figure 5), where the dyads were brought in from their home enclosures to the experimental arena via pipes they could walk through. The experimental arena consists of four compartments (60cm x 50cm x 50 cm each). The central compartments were separated by a transparent, removable partition, whereas the side compartments were separated from the central ones with a transparent partition and a small sliding door that could be manipulated to keep the individual within a particular compartment. Two platforms made of logs were placed in the central compartments where the individuals would sit to do the task. A hanging log and a hollow pipe were placed in each of the side compartments as enrichments for individuals to interact with before and after the task. When the sliding doors were open and the middle partition removed, the individuals could move about freely and interact with each other, as well as the enrichments within the experimental arena.

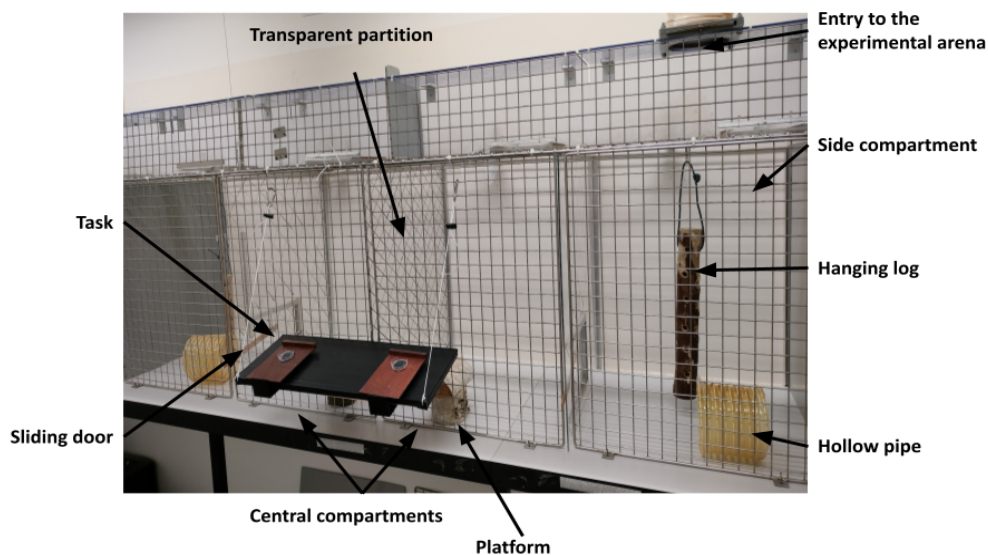


Figure 5. Experimental arena

Video and audio recordings

The primary video recording of the marmosets during the experiment was recorded using four Raspberry Pis (Raspberry Pi 4 Model B 2018, 4 GB RAM installed with Raspberry Pi OS 32-bit) connected to Raspberry Pi High-Quality camera modules (model V1.0 2018) and lenses (CCTV Lens 3 MP, 6 mm). The Raspberry Pi OS (32-bit) was installed on a SanDisk Extreme Pro 64 GB micro-SD card. The cameras were mounted on the wall using 3D printed stands, and Small Rig Ballhead camera stands. Each pair of cameras was positioned equidistantly from their corresponding half of the experimental arena, in a stereo-configuration at less than 90 degrees angle from one another, capturing each half from different angles. The lenses were focused and zoomed in to capture half of the experimental arena, as shown in Figure 6b.

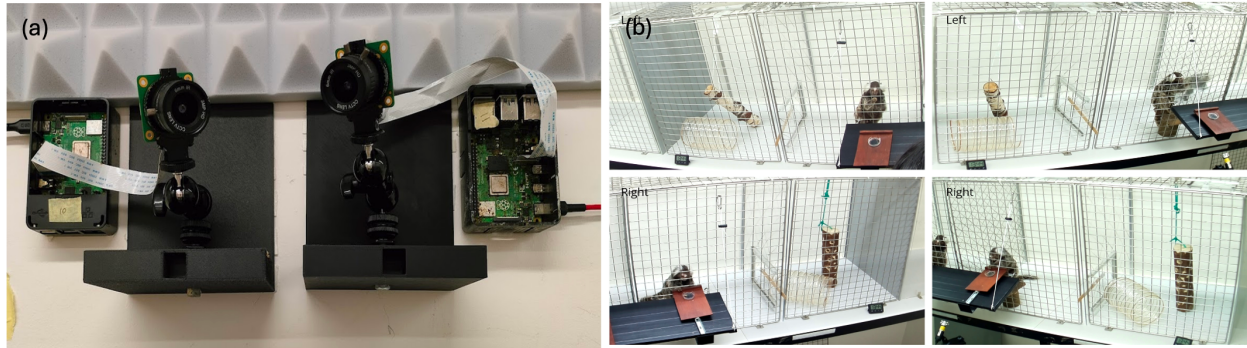


Figure 6. Orientation of Raspberry Pi cameras in the experimental room. a) Raspberry Pi cameras mounted on the wall, b) Field of view of the Raspberry Pi cameras

The video recording software on the Raspberry Pis and the main computer was a Python-based video recording program with a GUI that was developed by Dr Markus Marks when he was a postdoctoral researcher at the Neurotechnology Group, ETH Zurich. The videos were recorded in ‘*Long Acquisition*’ format, where small video segments of 2 minutes for the duration of the experiment (~15 mins) are recorded and saved, with a gap of 15 seconds between consecutive videos. The videos had a resolution of 1080 x 1920 pixels and were recorded at 24 frames per second.

An overview of the entire experiment, including the marmosets, the task and the experimenter, was recorded using a Sony Handycam HDR-CX200 set up on a tripod in the corner of the experiment room. This camera recorded video and audio and was used to code the number of prosocial pulls and the aggressive interactions in the session.

The vocalisations during the experiment were recorded using a condenser microphone (Avisoft-Bioacoustics CM16/CMPA) connected to a recording interface (Avisoft UltraSoundGate 116H), which, in turn, was connected to a laptop to start and save audio recordings using the Avisoft Recorder USGH software.

Experimental task

In order to study behavioural synchrony in marmoset dyads, we devised an experiment with a prosocial task and two minutes before and after the task where the marmosets are unrestricted and can interact freely within the arena. We had additional individual and joint reward tasks to contrast with the prosocial task, each lasting 3 minutes. The different tasks are described below. Prior to the start of the experimental session, each dyad went through ~10 trial sessions with the individual task to habituate them to the task and the waiting period in the experiment room.

Prosocial task

The prosocial task is a long sliding board with a detachable handle on one end, on top of a base wooden board connected by rolling sliders (Figure 7a). The sliding board has two food bowls on either end where food rewards can be placed. This apparatus can be hung outside a cage with hooks on the baseboard and suspended from carabiner hooks and cords. The board is hung at an angle such that the sliding board rolls back automatically and must be pulled by the marmosets to get the food reward.

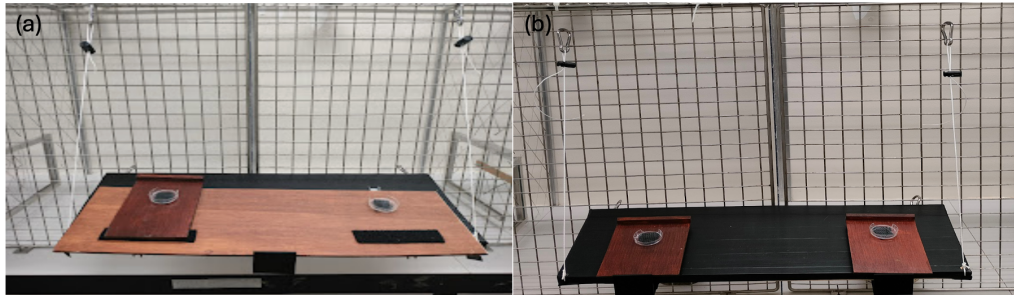


Figure 7. Prosocial and Individual sliding boards. a) Prosocial sliding board, b) Individual sliding board

A prosocial trial consisted of placing a food reward on the end of the board without the handle and no reward on the side with the handle such that when the focal individual pulls the board, the other individual receives the food reward. To execute a successful prosocial trial, both individuals must sit on the logs and wait in front of the task, and the focal individual must pull and hold the board until the receiving individual picks up the food reward. Each trial lasts 30 seconds after the food reward is placed in the bowl.

A prosocial session consisted of two alternating motivation trials and four alternating prosocial trials, with the handle switching between the two individuals. In each session, there can be a maximum of four successful prosocial trials. In the motivation trial, food rewards were placed in front of both the focal and receiver individuals to motivate the focal individual to pull the board. The prosocial session consisted of a total of six trials that went on for 30 seconds each, so the whole session lasted for about three minutes.

Individual task

The individual task consisted of two separate, smaller sliding boards attached to a base wooden board with food cups on top (Figure 7b). The separate sliding boards are also connected to the base with rolling sliders such that the boards slide back out of reach when the apparatus is hung from the cage at an angle. An individual trial involved placing food rewards on the cups at the same time so that both individuals got the reward simultaneously without having to coordinate with each other. Each trial lasted 30 seconds, and each individual session was composed of six individual trials, lasting for three minutes in total. Videos of the marmosets participating in prosocial and individual tasks can be viewed at this link - <https://tinyurl.com/38y8de5h>.

Joint reward task

The joint reward task was a modification of the prosocial task and was conducted using the prosocial sliding board, but in this case, both individuals got food rewards, but only one individual could pull the board. The detachable handle alternated back and forth between individuals, and they took turns pulling the board and obtaining food rewards for both. Like the other tasks, each joint reward trial lasted for 30 seconds, and six trials made up a joint reward session, where each individual got three chances to pull the long sliding board.

No-partner control session

The no-partner control session was designed to check if the marmoset dyads were pulling the board prosocially or if they were pulling the board due to a lack of inhibition when they saw a food reward. During this session, one individual was first restricted to the play compartment adjacent to the task compartment, where the focal individual would do the no-partner control task. The task consisted of six alternating self-reward trials and prosocial trials lasting 30 seconds each. During the prosocial trials, there is no individual on the other side to receive the food reward (Figure 8 bottom panel). If the marmosets understood the prosocial task, they would wait during the prosocial trials because there was no individual on the other side to receive the food. Once the task was completed for one individual, they would swap places, and the second individual would complete the trials while the first one was restricted to the play compartment. We compared the proportion of no-partner prosocial trials in which they pulled the board to the proportion of successful trials in the prosocial sessions. The control session takes about 12 minutes and there is no baseline or post-task period in the session.

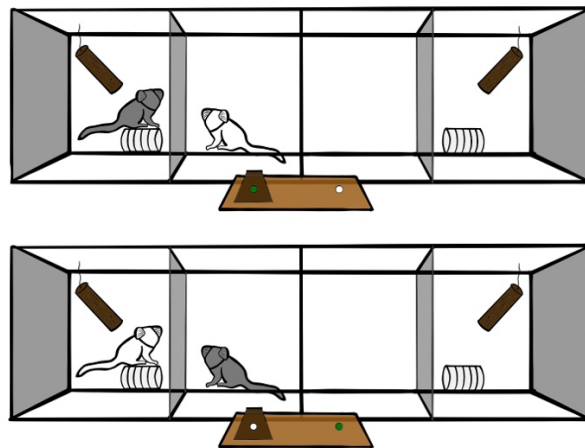


Figure 8. Illustration of the No Partner control task. Focal individual can pull in the top panel, but must not pull in the bottom panel, when the food (indicated by green circle) is in front of an empty compartment.

Each marmoset dyad participated in a no-partner control session, followed by three alternating prosocial and individual sessions each and another no-partner control session. This set was followed by three joint reward sessions (Figure 9b). The experimental sessions were conducted on consecutive or alternating days, based on the motivation level of the individuals.

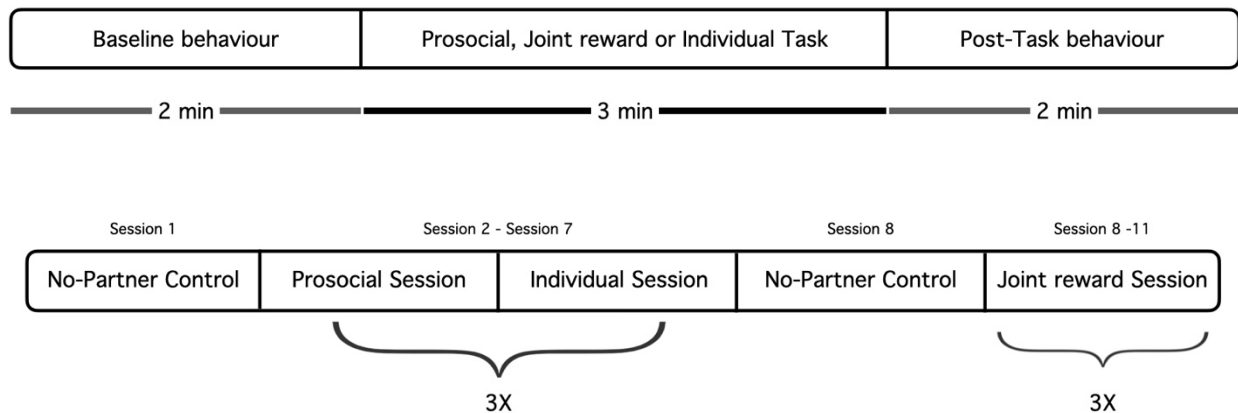


Figure 9. Timeline and order of the experiment sessions. a) Timeline of each experimental session, b) Order of experimental sessions for each dyad

Prosocial consolidation sessions

In addition to the trial sessions before the experimental sessions, we conducted prosocial consolidation sessions in the home enclosures to allow the marmosets to understand the task better. This involved restricting the dyads in a small cage-in-a-cage inside the home enclosure and presenting them with the prosocial task. The consolidation task involved two motivation trials followed by six prosocial trials, and this set was repeated twice. Ten consolidation sessions were conducted across two weeks. For three groups, these consolidation trials were conducted after the first set of experiment sessions, but before the joint reward sessions, and for one group (Nikitas), the consolidation trials were conducted before the start of the experimental sessions.

Behavioural coding

Prosocial pulls

The number of successful prosocial trials was counted from the overview recordings of experimental sessions. A prosocial trial was deemed successful if the focal individual pulled the board and held it in place long enough for the receiving individual to take the food reward from the bowl. For the no-partner control sessions, we counted the number of prosocial trials where the individuals pulled the board when there was no individual on the other side.

Task execution frames

We also manually coded the start and end frame numbers of the video segment where the individual is pulling the sliding board. The starting frame was where the individual lifted his/her hand to pull the board, and the end frame was when the individual picked up the food treat and brought it to his/her mouth.

Prosocial trial frames

We manually coded the start and end frames of each prosocial trial in the prosocial session, along with the ID and status of the individual, i.e., receiver or puller. The starting frame was when food was placed on the food plate and the ending frame was when the experimenter's hand came into view to switch the handle.

Video processing

The relevant video segments from each of the four Raspberry Pi cameras were merged and then spliced into three sets of videos for pre-task, task and post-task segments using FFmpeg, an open-source software to handle multimedia files. The task videos started from when both individuals were seated in front of the task in the central compartments to when the task and the partitions were removed. The task videos were cropped to half the width of the original video, since the individuals were restricted to the central compartments while they participated in the task. Additionally, these videos contained a single individual performing the task, since each camera focuses on one-half of the experiment arena. These task videos were further analysed for my master's thesis.

DeepLabCut projects

We created multi-animal DeepLabCut (maDLC) projects for each Raspberry Pi camera to track the poses of marmosets from different angles (Lauer *et al.*, 2022). First, we compared different models of pre-trained networks by training the models on 100 labelled images (of 7 body parts for two individuals) for 20,000 iterations with 95:5 train test split. The models were trained on a computer GPU (Intel UHD Graphics 630). The models were evaluated through train and test error, which gives us the mean reprojection error of training and test data in pixels, and we chose to go ahead with ResNet-152.

We decided to label nine body parts in freely moving marmosets – ears, nose, hands, legs, tail base and tail end. The ears were marked at the point where the ear tufts meet the top of the head. The ends of the forelimbs and hindlimbs were labelled as hand and legs, and the midpoint of the base of the tail was marked as tail base. The positions were labelled only if they were clearly visible and not occluded by anything else. To test for the accuracy of labelled body parts, we calculated the distance between the label positions of all body parts in 50 frames marked by two individual observers and compared the mean of the distance between the labels to its real-world dimensions.

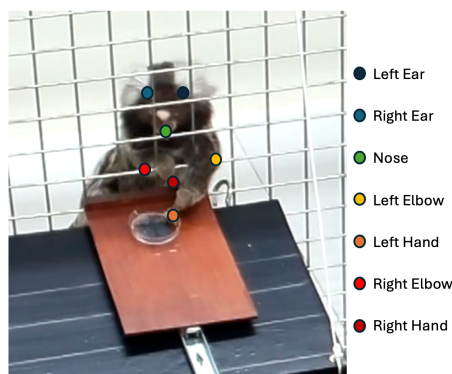


Figure 10. Body part labels on a marmoset for DLC training data. The dots indicate the labels for corresponding body parts marked manually to train DLC model.

For the task videos, marmosets were mostly sitting on the logs, with the sliding board occluding the lower half of the body. So, we omitted the legs, tail base and tail end labels and instead added elbows along with ears, nose and hands for each individual (as shown in Figure 10). In each maDLC project, a ResNet-152 network was trained with seven body part labels for 300 frames extracted from 27 videos through the k-means clustering algorithm. The network was trained for 50,000 iterations with a batch size of 2 and 95:5

train test split. It was then evaluated to confirm that the test error was less than 10 pixels (which corresponds to ~ 0.7 cm in real life dimensions at that resolution and distance from the camera). The trained network was subsequently used to analyse videos with identity set to False, and the detected tracklets were stitched together and filtered to create a smooth trajectory and saved as a CSV file for further processing.

Pose3D: Camera calibration

Pose3D is a semi-automated workflow that allows us to convert two or more 2D DLC detections into 3D coordinates (Sheshadri *et al.*, 2020). It is based on MATLAB StereoCameraCalibration included in MATLAB2022b Computer Vision toolbox (Zhang, 2000; Bouget, 2022). The process of camera calibration involves estimating certain intrinsic and extrinsic parameters of the cameras by detecting the corners in a checkerboard with a fixed number of squares of known dimensions (Figure 11). These parameters are then used to convert two sets of 2D coordinates into a 3D coordinates.

The intrinsic parameters of the camera are the lens and sensor properties – the focal length, principal point, radial and tangential distortion and skew of the cameras. The extrinsic properties describe the relative position of the cameras with respect to one another i.e., the translation and rotation matrices of one camera with respect to the other. These parameters are estimated automatically from extracting the position of checkerboard patterns in several pairs of images (Zhang, 2000). We used a checkerboard with 7 x 4 corners and each square was 35 mm in length. Two pairs of cameras were calibrated, one pair for the right and the other for the left half of the experimental arena. So we got two sets of intrinsic and extrinsic parameters.

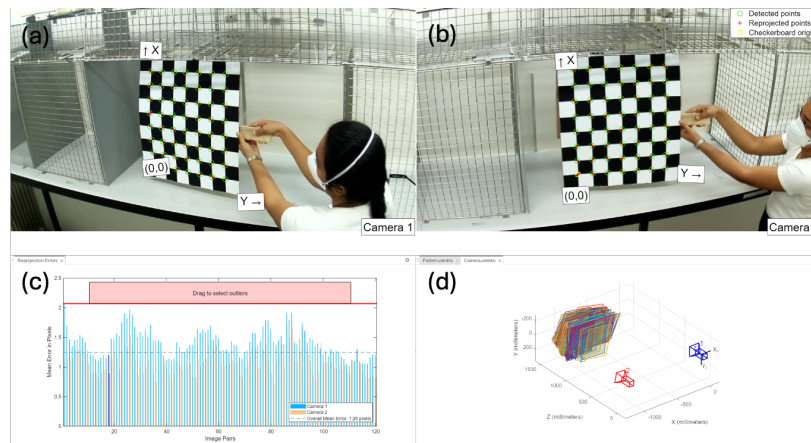


Figure 11. Camera calibration process in MATLAB. a) and b) The top panel contains a pair of checkerboard images which are used to estimate camera parameters, and the bottom panel contains the results of calibration: c) mean reprojection error and d) a visualisation of the position of cameras.

Intrinsic parameters

The algorithm for the camera calibration is as follows. It assumes a pinhole camera model and for a given camera, there exists a transformation that relates 2D coordinates to real world 3D coordinates in the following way –

$$w \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K [R \quad t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Here, (x, y) is the image coordinate and (X, Y, Z) is the real-world coordinate. K is the intrinsic matrix of the camera, R is the 3D rotation matrix and, t is the translation matrix of the camera, and w represents the scale factor. The intrinsic matrix is further defined by –

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

In this matrix, the coordinate (c_x, c_y) represents the principal point (the point on the image plane where the perspective center is projected) and s is the skew, which is 0 if the X and Y axes are perfectly perpendicular.

$$f_x = F * s_x$$

$$f_y = F * s_y$$

F is the focal length of the camera, calculated in real world units and (s_x, s_y) are the number of pixels in world unit along the X and Y axes respectively. Once the intrinsic parameters are estimated, they are first used to undistort the 2D coordinates.

Extrinsic parameters

The extrinsic parameters describe the external 3D geometry of the two cameras. The relative pose (translation and rotation) of the cameras is described as follows where *Orientation* and *Location* describe the absolute pose of the camera, R is the rotation matrix and t is the translation matrix.

$$\text{Orientation1} = \text{Orientation2} * R$$

$$\text{Location1} = \text{Orientation2} * t + \text{Location2}$$

Essential (E) and Fundamental (F) matrices are parameters used to transform 2D coordinates to 3D. They are described by the geometric relationship between the corresponding pairs of coordinates from each camera through the following constraint –

$$\begin{bmatrix} p_1 \\ 1 \end{bmatrix} * F * \begin{bmatrix} p_2 & 1 \end{bmatrix} = 0,$$

where p_1, p_2 are points in image 1 and 2 expressed in pixel coordinates

$$\begin{bmatrix} P_1 \\ 1 \end{bmatrix} * E * \begin{bmatrix} P_2 & 1 \end{bmatrix} = 0,$$

where P_1, P_2 are normalized points in image 1 and 2, where the origin in the optical center of the camera and the coordinates are scaled by f_x and f_y

The calibration returns these parameters and the mean reprojection error in pixels. Pose3D takes in DLC detections as input and uses MATLAB's camera calibration to undistort and triangulate 2D coordinates and

return 3D coordinates from multiple cameras. The semi-automated code allows us to change parameters to choose 2 best coordinates out of multiple detections, set a likelihood threshold (such that the detection below that will be excluded) and further filter and plot 3D reconstructed trajectories. Analysed videos with detections and corresponding 3D reconstructed videos can be viewed at this link - <https://tinyurl.com/38y8de5h>.

Processing DLC detections

Pre-processing for 3D alignment

Analysing videos in DLC returns the 2D coordinates of body parts and the associated likelihood value (ranging from 0 to 1) of the detection in a CSV file. When a body part is occluded or the individual goes out of view, the detection and likelihood values are also omitted. However, the MATLAB calibration algorithm does not accept Not a Number (NaN) values, which means that there are missing detections that need to be interpolated before 3D reconstruction. We first quantified the missing detections by calculating the percentage of missing detections, the maximum number of frames for which a body part or the whole individual goes missing and the distribution of median likelihood values across all detections.

We then linearly interpolated the missing coordinates, but not the corresponding likelihood values, which would remain NaN. During 3D reconstruction using Pose3D, we set the likelihood threshold parameter to 0.8 (based on the distribution of median likelihood values), which removed any 3D point where either of the cameras had a detection with likelihood value less than 0.8.

3D Transformation

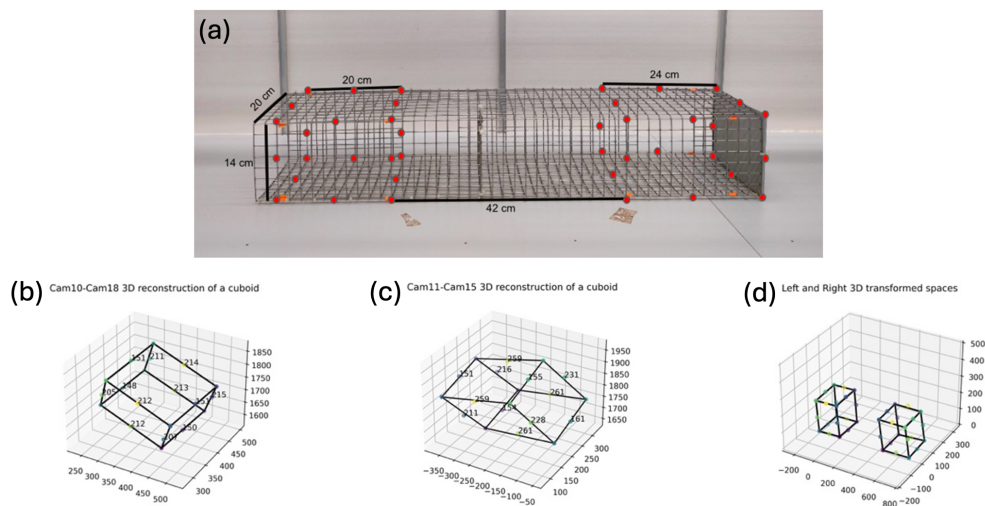


Figure 12. 3D reconstruction and transformation of cuboids for 3D alignment. a) Long cuboid with known dimensions placed such that it's visible from all cameras, b) Cuboid reconstruction on left half with edge lengths, c) Cuboid reconstruction on right half with edge lengths, d) Transformed cuboids aligned to the real-world axes.

Once we had the reconstructed 3D coordinates from the left and right half of the arena (Figure 12b and 12c), we had to transform them to align them to the real-world coordinates (Figure 12d). We annotated the corners and edges of similar cuboids in the left and right half of the experimental arena. The cuboids were of known dimensions at a particular distance from each other as can be seen in Figure 12a. First, we confirmed that the

opposite edge lengths of the cuboid were equal to make sure that the 3D transformation was linear (Table 1). We then chose one of the corners as a reference point and translated the whole space such that this corner was the origin. Once the cuboid corner was centred, we serially calculated the angles between the edges of the cuboid and the adjacent axis such that when rotated, the cuboid edges would align with the axes (Table 2). Next, we scaled the right 3D space to match the coordinates in the left 3D space (Table 3). Finally, the right 3D space was translated by 440.46 units (estimated by using the ratio between the edge length in 3D reconstructed space and its real-world length) such that the cuboids were placed in the same 3D space as in the real world (Figure 12d). All 3D reconstructions of the marmoset body parts were translated, rotated and scaled before further analysis.

Table 1. Edge lengths of the cuboids in left and right 3D reconstructed space and real-world space

Cuboid	Edge 1	Edge 2	Edge 3
Left (cm)	14	20	20
Left_3d (pixel, stdev)	149.96, 1.07	209.36, 3.66	212.48, 0.71
Right (cm)	14	20	24
Right_3d (pixel, stdev)	155.20, 3.60	221.38, 8.22	260.15, 0.88

Table 2. Serial rotations of the 3D coordinates to align the cuboids with real world axes

Left 3D space		Right 3D space	
Angles	Axes	Angles	Axes
-0.396	y	0.307	y
-0.224	x	-0.307	x
0.067	z	0.143	z
-1.57	x	-1.57	x

Table 3. Scaling factors of the right 3D space with respect to the left 3D space

X axis	0.98
Y axis	0.97
Z axis	0.95

Processing 3D detections for gaze analysis

The direction of the head was extracted by calculating a vector from the midpoint between the ears to the nose of the marmosets using a function written in Python. I first selected the 3D coordinates of ears and nose for each individual and removed segments if all the detections were missing for 30 or more consecutive rows i.e., if the animal moved out of frame for a period of time. This was based on the distribution of the maximum number of rows the entire individual is missing across all videos. Following this, the 3D data was

interpolated using the cubic spline method and the head vectors of the individuals were calculated for every frame.

We then calculated the angle between the head vectors at every instance by taking the cos inverse of the dot product of the vectors divided by the product of the magnitude of the vectors as described in the equation below. In order to differentiate between the head vectors pointing away versus pointing towards each other at the same angle, we looked at the sign of the dot product between the Z axis and the cross product of the two vectors (Figure 13).

$$\alpha = \cos^{-1} \left(\frac{\vec{A} \cdot \vec{B}}{|\vec{A}| \times |\vec{B}|} \right)$$

$$\text{If } \vec{Z} \cdot (\vec{A} \times \vec{B}) > 0, \quad \theta = \alpha$$

$$\text{If } \vec{Z} \cdot (\vec{A} \times \vec{B}) < 0, \quad \theta = \pi - \alpha$$

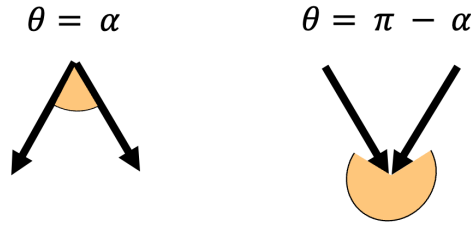


Figure 13. Illustration of calculation of angle between head vectors. The arrows indicate the head vectors (proxy of gaze direction) and the yellow shaded region indicates the angle between head vectors of individual 1 and individual 2.

In order to calculate the fraction of time for which marmosets look at their own food board versus the board of the other individual, we manually extracted the 3D coordinates of the corners of the sliding board and divided it in half to assign each half to the corresponding individual. We then used three coordinates on the board to define a plane and checked if a line extended by the head vector intersects the plane or not. If it intersects the plane, we then calculated the point of intersection on the plane and checked if it's projection on the edges falls within the boundaries of the board. The relevant equations are listed below and illustrated in Figure 14.

$$\vec{X} = P3 - P2, \quad \vec{Y} = P1 - P2, \quad \text{where } P1, P2, P3 \text{ are points on the plane}$$

$$d = \frac{(P2 - T) \cdot (\vec{X} \times \vec{Y})}{\vec{A} \cdot (\vec{X} \times \vec{Y})}, \quad \text{where } T \text{ is the midpoint of ears and } \vec{A} \text{ is the headvector}$$

$$\text{If } d > 0, \quad P = T + \vec{A} \cdot d, \quad \text{where } d \text{ is the distance from the } T \text{ to } P, \text{ the point of intersection on the plane}$$

$$\vec{P} = P - P2, \quad \vec{P}_x = P2 + \frac{\vec{P} \cdot \vec{X}}{\vec{X} \cdot \vec{X}} \vec{X}, \quad \vec{P}_y = P2 + \frac{\vec{P} \cdot \vec{Y}}{\vec{Y} \cdot \vec{Y}} \vec{Y}, \quad \text{where } \vec{P} \text{ is a vector of } P \text{ from } P2$$

If $|\vec{P}_x| < |\vec{X}|$ **and** $|\vec{P}_y| < |\vec{Y}|$ then the extended head vector intersects the food board

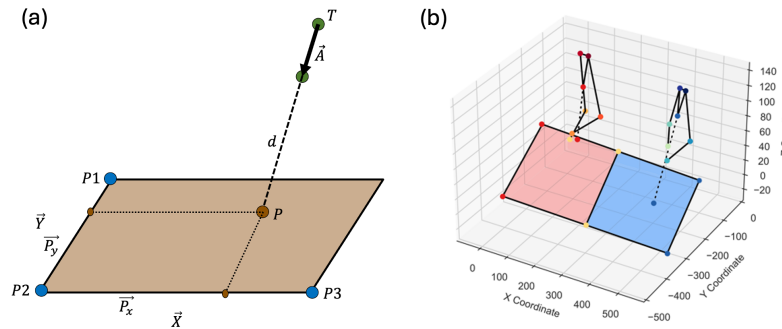


Figure 14. Illustration of gaze intersection with sliding board. a) Illustration of calculation of line-plane intersection using three points on the plane and the head vector b) Illustration of the intersection of head vectors of two marmosets with the self or other half of the sliding board

Thus, using the coordinates of the points on the sliding board, the nose and the midpoint between the ears of the marmosets, we can check if an individual's head vector intersects with its own part of the sliding board or the other half of the sliding board in every frame. We verified the 3D reconstruction and transformation by creating a video of the detections of the two individuals (with the skeleton), the task board and the gaze intersection point for a small clip and comparing it to the overview video.

Often, the head vector had a lower elevation (angle with the Z axis) than the actual direction of the gaze. In order to correct for the elevation, we manually labelled the ears, nose and food reward for each individual, where they were clearly looking at the food and calculated the angle between head vector and the vector between the food and midpoint between ears, called the gaze vector. The difference in angle between the head and gaze vectors ranged from 15° to 35° and the average elevation was 25° . For each frame, we computed a range of gaze vectors (elevated by 15° , 25° and 35° along the Z axis) and considered the gaze intersection with the board to be True if any of the elevated gaze vectors intersected the plane. We then calculated the proportion of time for which the individuals' gaze intersected their own board and the other board for each video and compared it across different conditions.

We also analysed the gaze-board intersection during the Prosocial sessions in detail by assigning the role of 'Puller' or 'Receiver' to the individuals and differentiating the time spent looking at the other individual's board by the status of the focal individual. We also compared the time the receiver and puller spent looking at the task board when it was a successful prosocial trial versus an unsuccessful one. All statistical tests were done using SciPy and Scikit-posthocs libraries in Python, and linear mixed-effects models were fit using lmerTest package in R.

Processing 3D detections for task kinematics analysis

In order to analyse the similarities and differences in the execution of the task, we focused on the individual condition and selected the segments of the 3D coordinates where the marmosets are pulling the sliding board, using the manually annotated start and end frame for each pull. These segments were then interpolated using the cubic spline method before analysing the detections further. We first looked at the absolute value of the log of the ratio of path lengths of the right- and left-hand detections to check if they use one hand more

dominantly than the other. If the value is close to 0, that means they use both hands equally in pulling the board, and if it's much greater than 0, they use one hand more dominantly than the other.

$$handedness = abs\left(\log\left(\frac{left_hand_path}{right_hand_path}\right)\right)$$

We used dynamic time warping to analyse the similarity in the movement of the arms across individuals. First, we calculated the DTW distance using the 'dtw_ndim' function from the **dtaidistance** Python library for both elbows and both hands 3D trajectories for the duration when the individual was pulling the sliding board (Meert *et al.*, 2020). Before analysing the trajectories, we translated the coordinates such that the individual sliding boards were superimposed such that the two individuals overlapped as much as possible. First, to contrast the intra-individual and inter-individual differences in a dyad, we compared DTW distance of the pulls of one individual with its own pulls in the same session (intra-individual) with the DTW distance of the pulls of two individuals of a dyad in the same session (inter-individual).

Then, we compared the distance metric between the pulls of two individuals in real dyads in a session with pseudo dyads to ensure that the similarity of the trajectories was above chance levels. Since three dyads out of six were from the same group, we split them into two sets – one with dyads from the Nikitas group and the other set containing the other three dyads (Jupie-Mercur, Lancia-Lexus and Vesta-Vito). We constructed pseudo dyads by combining pulls from a session of the individuals from set 1 with those of set 2. This ensured that a pseudo dyad wouldn't be created from the same individual from different dyads or sessions.

The code will be made available upon request.

Chapter 3: Results

Prosocial task results

The following plots show the prosocial rate for each dyad across three sessions – calculated by dividing the number of prosocial pulls by the maximum possible prosocial pulls in a session (Figure 15). Two dyads out of six had a positive prosocial rate in at least 2/3 sessions (driven by two individuals), and two other dyads had a prosocial rate of 0 across three sessions. In one dyad (JM), the female was food-dominant and made aggressive chatters towards the male during some prosocial sessions.

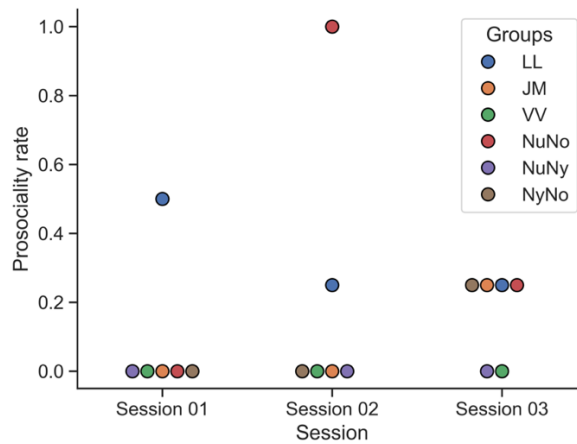


Figure 15. Rate of prosocial pulls across prosocial sessions. Different colours represent six different dyads

All individuals, except 2 (Lancia and Mercur), perform well in the no partner control task (Figure 16), starting off with a low pulling rate in session 1 and decreasing it further in session 8, which indicates that they're pulling prosocially in the experiment and not due to of lack of inhibition.

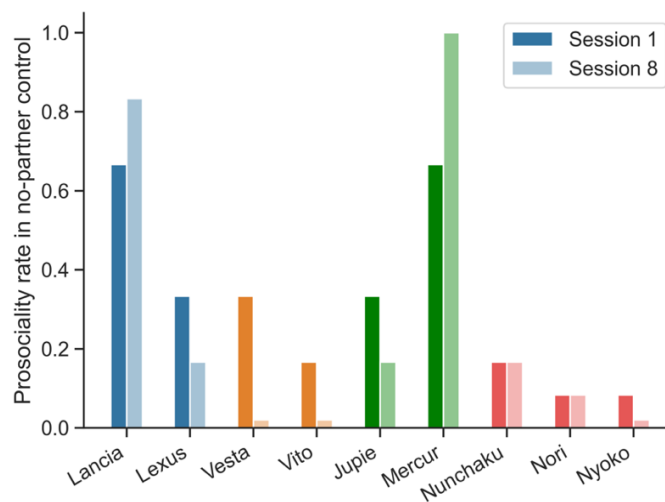


Figure 16. Rate of pulls of the prosocial board in no-partner control session when there is no individual on the other end to receive food. The different colours represent individuals belonging to different groups.

Figure 17 shows the number of prosocial pulls out of a maximum of three pulls by each individual in the consolidation sessions. All individuals, except Nunchaku, didn't pull prosocially in at least half the sessions. The consolidation sessions didn't improve the performance of the dyads even after 10 days of trials.

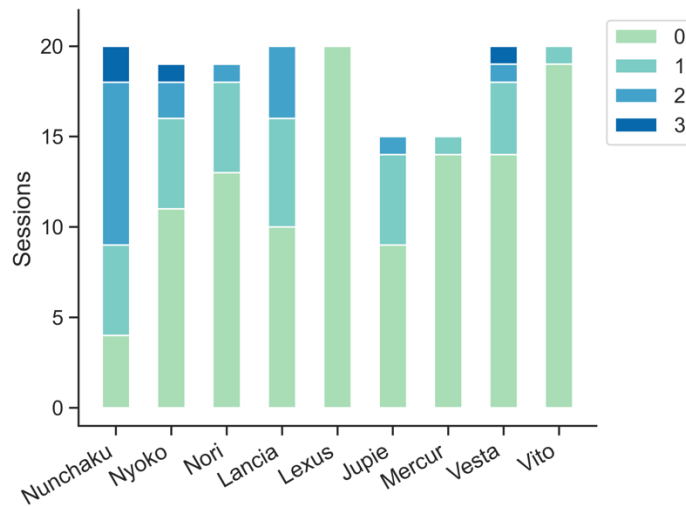


Figure 17. The number of prosocial pulls in the prosocial consolidation sessions. The maximum possible pulls in each session is 3, and 20 sessions were conducted for each dyad.

DeepLabCut model comparison and evaluation

DLC is based on transfer learning, which means there are several pre-trained neural networks from which to choose. We first wanted to compare the performance of different models on our dataset before using the model to train and analyse marmoset videos. Based on the train and test error of the models (given in pixels), ResNet-152 performed the best (Figure 18a). All the following DLC projects were trained on this network.

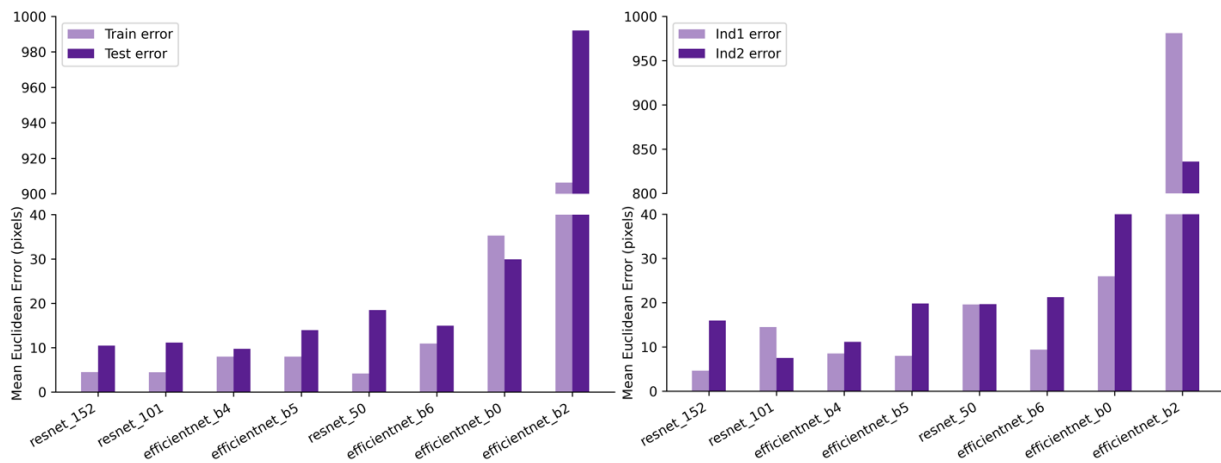


Figure 18. Comparing the performance of DLC models. a) Train and test error, b) Individual error of different DLC models when trained on the same data set of 100 labelled frames for 20,000 iterations

Next, we wanted to validate the training dataset that we would be providing to the neural network. The mean of distance between labels across 9 body parts of two individuals was 5.56 and 5.44 pixels (as shown in Figure 19), which corresponds to ~ 0.38 cm in real world at that resolution of the video and the distance of the experimental arena from the cameras.

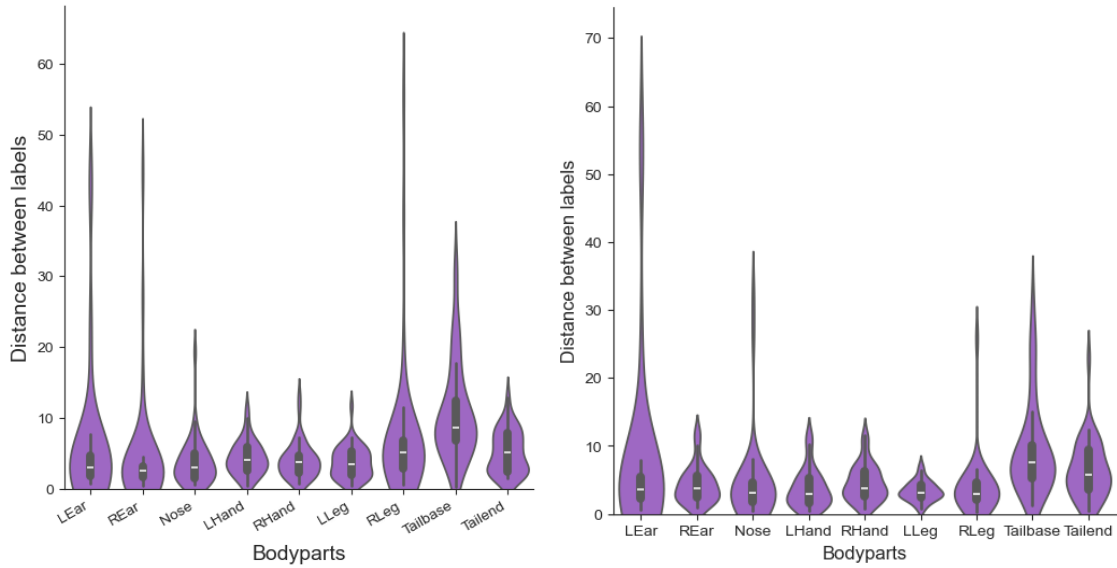


Figure 19. Accuracy of DLC labels. Accuracy of labelling of body parts of a) marmoset 1 and b) marmoset 2 for 50 frames using DeepLabCut

After training the DLC projects for each camera on 300 task frames for 50,000 iterations, we evaluated the model and ensured that the test error was less than 10 pixels. The summary of model evaluation from four task projects is plotted in Figure 20.

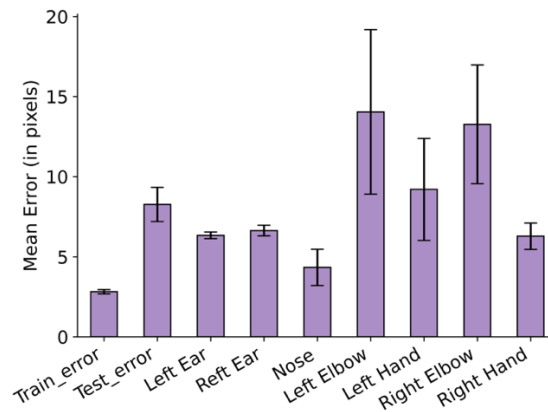


Figure 20. DLC model evaluation summary. Model evaluation of the four maDLC projects trained on 300 labelled frames for 50,000 iterations and used to analyse task videos

3D Camera Calibration

The camera intrinsic and extrinsic parameters estimated through 3D camera calibration are listed in detail in the Appendix section. The mean reprojection error, measured in pixels, is a measure of the accuracy of the estimation. The mean reprojection error for the left and right cameras were 1.47 and 0.69 pixels, respectively.

DLC Detections Summary

Once all the videos were analysed using DLC, we quantified the fraction of undetected body parts and plotted the histogram with mean and 95th percentile value of the percentage of detections linearly interpolated before 3D reconstruction (Figure 21). We also quantified the maximum number of continuous rows for which all the detections of the individual are missing in a video to decide the threshold for interpolation after 3D reconstruction, which was set to 30 rows (1.25 second) based on the early peaks (Figure 22). We also plotted a histogram of the mean of median likelihood values across all body parts of the individual in a video to decide on a threshold for 3D reconstruction, which was set to 0.8 based on the histogram (Figure 23) and previous studies which have used DLC (van der Zouwen *et al.*, 2021).

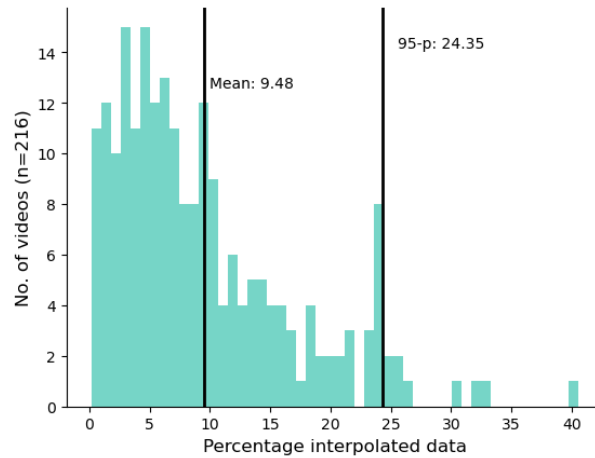


Figure 21. Distribution of percentage of missing detections across all body parts

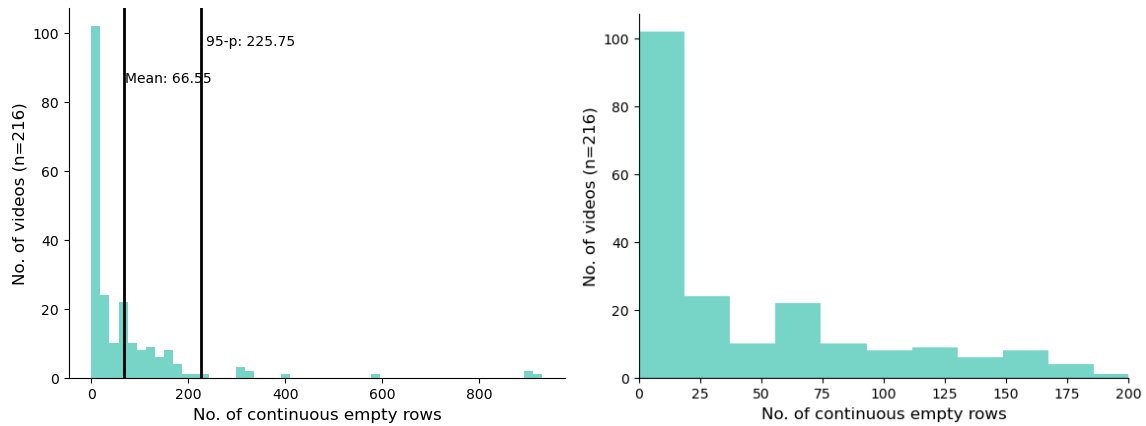


Figure 22. Distribution of the maximum number of continuous null detections a) Distribution of the number of continuous rows for which whole animal tracking is out of view, b) Magnified view of the distribution with a smaller x-axis range

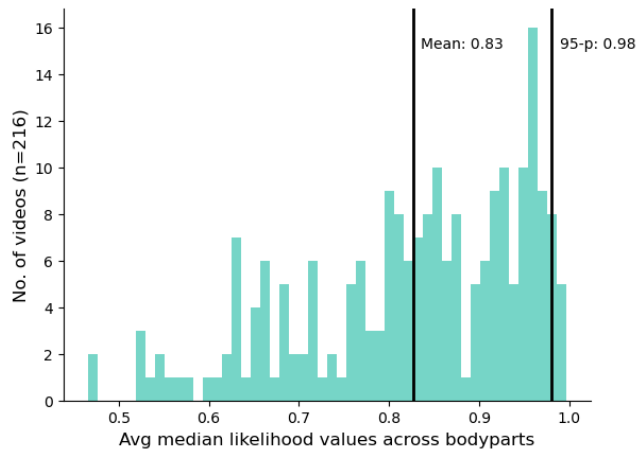


Figure 23. Distribution of median likelihood value across all body parts. This was used to decide the likelihood threshold for 3D reconstruction.

Gaze analysis during tasks

In order to study the use of mutual gazing during task participation, we first calculated the angle between the direction of their gaze given by the angle between the head vectors of the individuals. Figure 24 shows the distribution of angle between individuals across three conditions extracted from 18 videos, each video containing about 4000 data points. Irrespective of task condition, the angle between head vectors of individuals shows a bimodal distribution with two peaks – a higher peak at about 325° and a slightly smaller one at about 45°. A peak at 325° indicates that both individuals are looking in the same direction by an angle of 35°, and the peak at 45° means their head directions are aligned, looking away from one another at an angle of 45°. There are very few instances when the angle between the vectors is between 135° to 225°, that is when the individuals are looking in widely different directions.

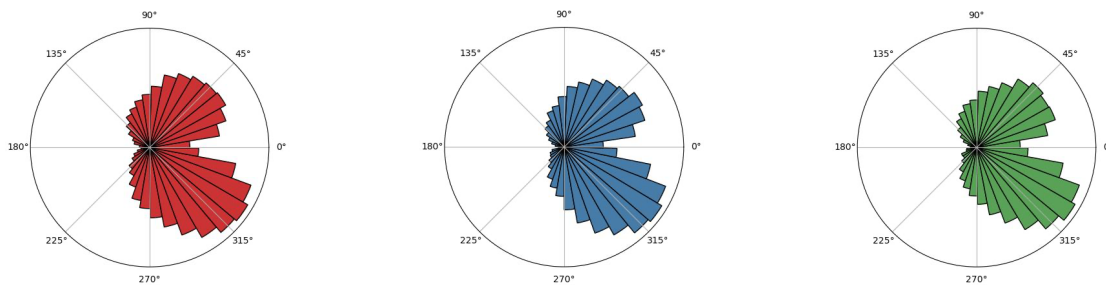


Figure 24. Distribution of angle between the head vectors of individuals in a) prosocial task, b) joint reward task, c) individual task. Each distribution contains about 4000 data points from 18 videos each.

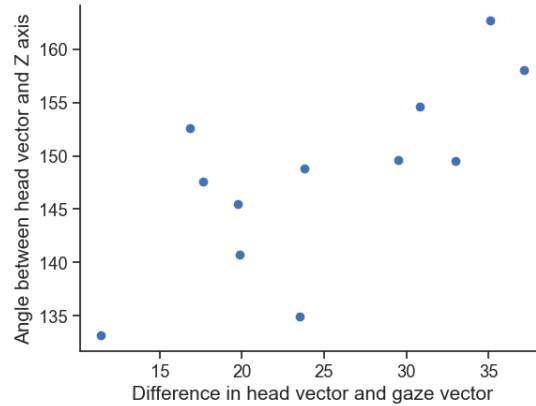


Figure 25. Estimation of elevation angle of the head vector. Distribution of the angle between the head vector and gaze vector (X-axis) and the corresponding elevation of the head vector with respect to real world Z axis (Y axis) calculated for 12 individuals from 12 different frames

We first validated the gaze detections by calculating the angle between the head vector and gaze vector of individuals, most of which fall within the range of 15°-35°, with an average of 25° (Figure 25). We then quantified the fraction of time during the task when the individuals’ gaze intersects their own food board versus the other individual’s food board across different task conditions. After checking for the normality and homoscedasticity of the gaze-self board intersection data (Figure 26a), we performed a one-way ANOVA, which showed that there’s a significant difference in the proportion of time spent looking at their own board across conditions (statistic: 3.588, p-value: 0.031). We then performed a post-hoc Tukey test and found a significant difference between the prosocial and individual conditions (p-value: 0.0235). We also fit a linear mixed model to the gaze intersection data using the lmerTest library in R. The models were fit as follows –

$$model1 = lmer(self \sim condition + (1 | group))$$

$$model2 = lmer(other \sim condition + (1 | group))$$

The gaze intersection time was the dependent variable, the condition was the independent variable, and the group ID was considered as the random effect. We then did an ANOVA on these linear mixed models, which were significant (model1 p-value: 0.0114; model2 p-value: 0.0068) and confirmed that the effect of the session conditions was significant on gaze-board intersection time even after accounting for variation across groups.

Values in the gaze-other board intersection time were not normally distributed (Individual – Shapiro-Wilkes score: 0.931, p-value: 0.026; Joint reward – Shapiro-Wilkes score: 0.920, p-value: 0.0133). So, we performed the Kruskal-Wallis test and found a significant difference between the proportion of time spent looking at the other individual’s board across conditions, as can be seen in Figure 26b (statistic: 9.286, p-value: 0.00962). We then did a post-hoc Dunn test with the Bonferroni correction and found a significant difference between Prosocial and Individual tasks (p-value: 0.00908).

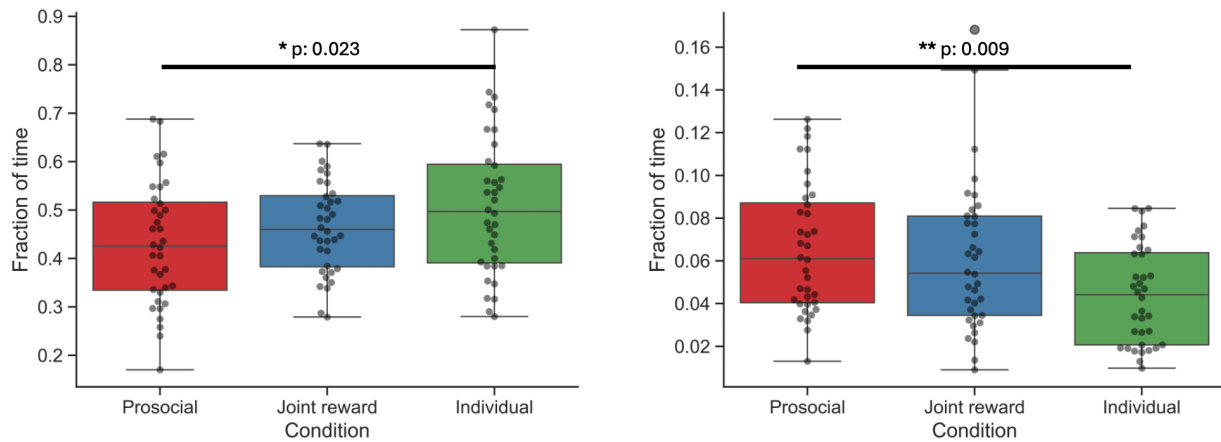


Figure 26. Fraction of time, the individual's gaze intersects the sliding board. a) Fraction of time the individual's gaze intersects with their own food board ($n=36$ for each condition, ANOVA [statistic: 3.588, p -value: 0.031] with posthoc Tukey test - Prosocial-Individual p -value: 0.0235), b) Fraction of time the individual's gaze intersects with the other's food board ($n=36$ for each condition, Kruskal-Wallis [statistic: 9.286, p -value: 0.00962] test followed by posthoc Dunn test with Bonferroni correction – Prosocial-Individual p -value: 0.00908)

We analysed the fraction of time spent looking at the task board during the prosocial trials ($n=72$), differentiating the individuals by their status as 'puller' or 'receiver' during the trials. We see that receivers spend more time looking at their own board and less time looking at the other individual's board as compared to the pullers (Figure 27a). We also compared the total fraction of time the individuals spend looking at the task board during successful ($n=11$) vs. unsuccessful ($n=61$) prosocial trials (Figure 27b). Both receivers (Welch's T-test – statistic: -2.281, p : 0.0189) and pullers (MannWhitneyU test – statistic: 213.0, p : 0.032) spend significantly more time looking at the board during successful trials as compared to the unsuccessful trials.

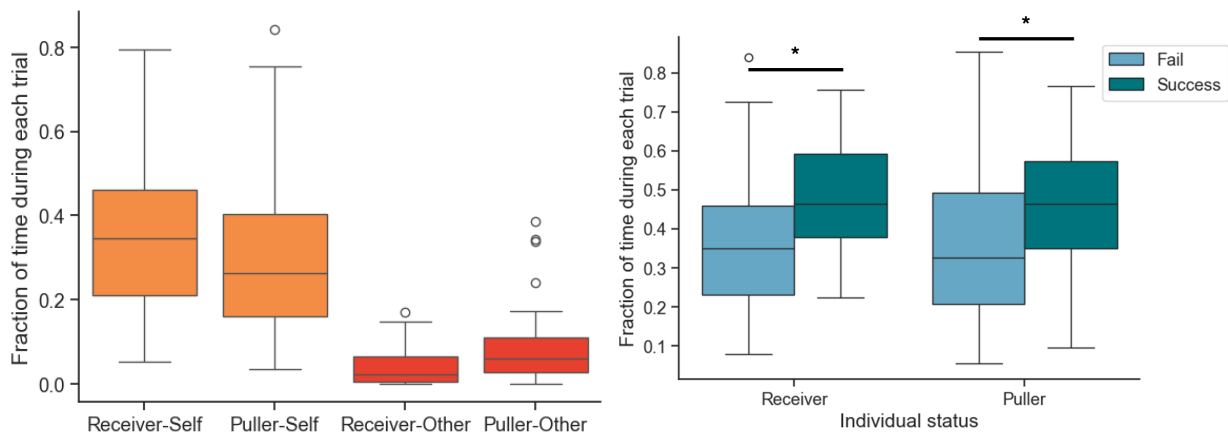


Figure 27. Gaze intersection during prosocial session differentiated by individual status. a) Fraction of time spent looking at its own board versus the other board differentiated by individual status (receiver or puller, $n_{total}=72$), and b) Fraction of time spent looking at the board based on success ($n_{success}=11$, $n_{failure}=61$) of the prosocial trial, differentiated individual status (receivers - Welch's T-test – statistic: -2.281, p : 0.0189; and pullers - MannWhitneyU test – statistic: 213.0, p : 0.032)

Task kinematics in the individual task

As a first step to analysing the similarity in task execution, we chose to focus on the kinematics of board pulling in the individual task. We calculated the path traced by the detections of the hands and elbows of the individuals and compared the absolute value of the log of the ratio of left-hand path length to the right-hand path length of the individuals to get a measure of their handedness across six trials in a session and across three sessions (Figures 28a, 28b). There is a lot of variability in difference in handedness across various trials and sessions, with no striking trend. We also looked at the time taken to pull the board in each trial, but we didn't observe any difference in the duration across trials or across sessions (Figures 28c, 28d).

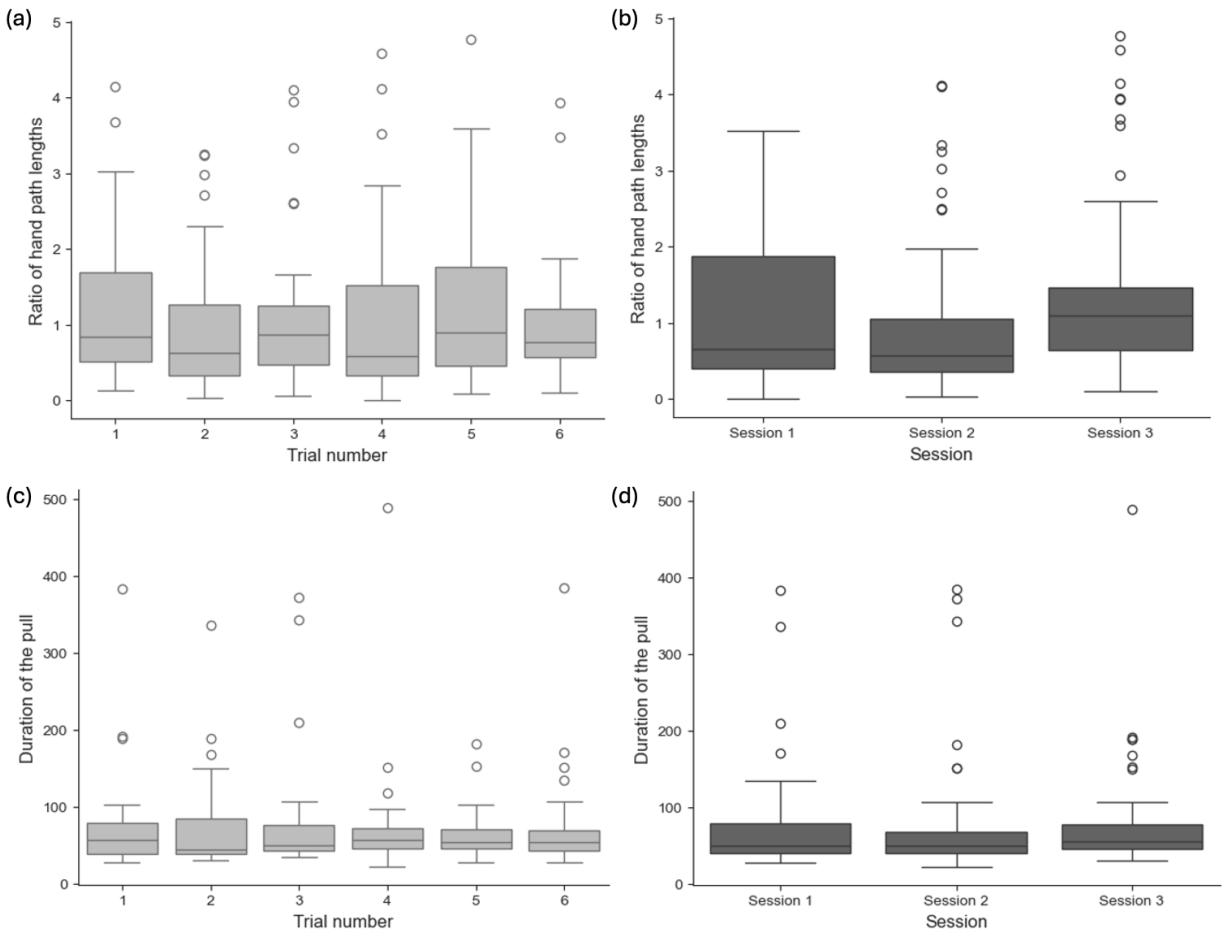


Figure 28. Measure of handedness and duration of task execution from hand detections. The difference in handedness of the pull by a) trials within a session, b) across three individual sessions; Duration of the pulling action, c) across trials within a session, d) across three individual sessions ($n_{\text{total}}=187$ trials)

When it comes to the similarity in execution of the task, first, we looked at intra-individual ($n=36$) and inter-individual ($n=18$) differences in a session (Figure 29a). Across both hands and elbows, we see that the trajectories of an individual's pulls are significantly more similar to themselves than to the trajectories of the other individual in the same session. Table 7 summarises the results of Mann Whitney U tests (testing against 'less' alternative hypothesis) performed on DTW distance values for each body part. The lesser the value of the distance measure, the more similar the trajectories are. We also compared the DTW distances between

trajectories of two individuals from a real dyad (n=18) in the same session with the distance metric of trajectories of pseudo-dyads (n=324), and we don't see a significant difference between the DTW distance values (Figure 29b, Table 7).

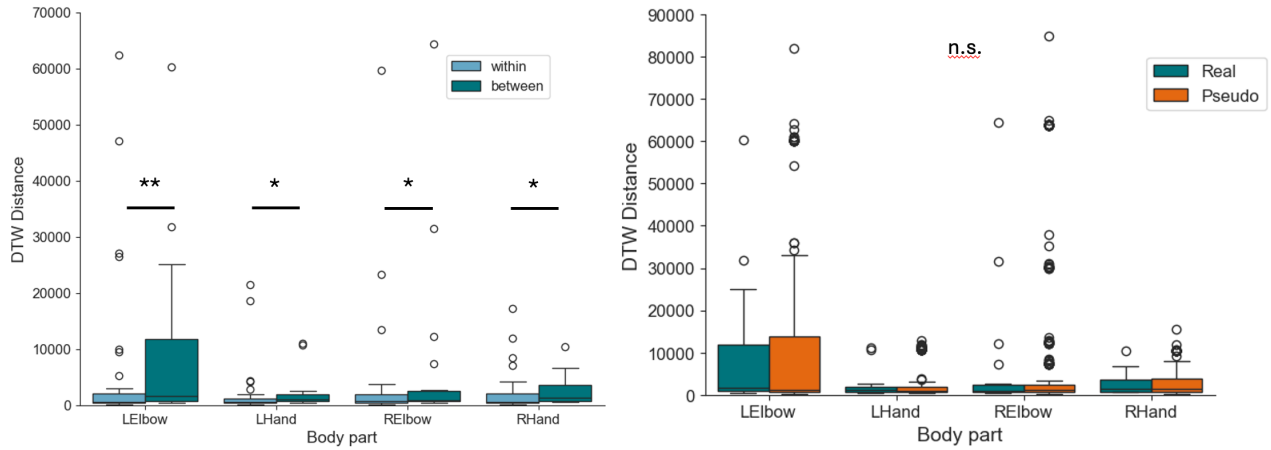


Figure 29. Dynamic time warping distance measure of task execution a) Comparing pulls within the same individual (n=36) and pulls between two individuals of a dyad in an individual session (n=18) with pairwise Mann-Whitney U test, b) Comparing pulls between real dyads (n=18) and pulls between pseudo dyads across individual test sessions (n=324)

Table 4. Statistical test results of DTW distance comparison. Test statistic and p-value for Mann-Whitney U Test for differences in DTW distances across intra- and inter-individual combinations, and real vs. pseudo dyads

	Intra- and Inter-individual differences		Real and Pseudo-dyad differences	
	Test statistic	p-value	Test statistic	p-value
Left elbow	190.0	0.0071 **	3238.0	0.4310
Left Hand	205.0	0.0148 *	3196.0	0.4936
Right Elbow	232.0	0.0465 *	2861.0	0.8938
Right Hand	199.0	0.0111 *	3074.0	0.6996

Chapter 4: Discussion

Common marmosets are cooperative breeders and engage in interdependent, coordinated tasks in their everyday interactions. Marmosets tend to tune in with each other and align themselves with other individuals in the group to improve coordination in infant care (Burkart *et al.*, 2009; Guerreiro Martins *et al.*, 2019). Our objective was to examine if marmosets, like humans, exhibit behavioural synchrony as a proximate mechanism for improving cooperation and coordination in the group. We designed an experiment to test the effects of prosociality on behavioural synchrony by having marmoset dyads interact freely before and after the prosocial task to compare the levels of behavioural synchrony and pose imitation. We also wanted to analyse gaze dynamics and task kinematics in marmoset dyads engaged in different task conditions. We implemented an automated pose tracking and 3D reconstruction pipeline, using DeepLabCut and Pose3D, to extract 3D trajectories of body parts of freely moving marmosets. We used this pipeline to analyse gaze direction and hand kinematics in marmoset dyads while they participated in different tasks. We observed distinct gaze patterns across different tasks, but no significant similarities in hand kinematics of a real dyad as compared to a pseudo dyad.

The marmoset dyads in my study were not highly prosocial towards their partner (Figure 15), although most individuals understood the task, based on the no partner control session (Figure 16). We also know from previous and ongoing studies that marmosets exhibit intentional prosociality (Burkart and van Schaik, 2020) and willingly provide food for the rest of the group members in a task termed as group service. This could be due to one or many of several reasons. One reason could be the group size and group composition. Three out of six dyads in our study live in dyadic groups, without infants or helpers, which decreases the drive to coordinate with the other individual. Four dyads were siblings (Nikitas and Vesta-Vito), and one of the other two breeder dyads (Jupie-Mercur) had never produced an offspring, which also adds to the lack of motivation for proactive prosociality in these dyads. Moreover, it could be that prosociality is more of a group level phenomenon with a strong influence of group size, due to a higher probability of coordination from at least one individual in the group.

In addition to this, the experimental room is an alien and sterile condition for marmosets, even after habituation, as compared to their home enclosure, which increases their arousal, and they are unwilling to participate in the task unless they get some food reward. The previous experiments on prosociality were conducted in a different experimental room (Burkart and van Schaik, 2020), and group service tasks were done in their home enclosures, where the marmosets were more at ease.

One of the major goals of the project was to set up an automated tracking software to analyse the movement and posture of freely interacting marmosets and thus provide a proof of concept of this approach. Previous efforts in the lab had attempted to achieve this with SIPEC, a deep learning based behaviour classification tool that segments videos, identifies individuals, and performs pose estimation and behaviour classification, but it did not work very well with marmosets when they were far away from the camera or partially occluded (Marks *et al.*, 2022). We decided to use DeepLabCut, which estimates the pose of the individuals from body parts of interest, allowing us to explore postural synchrony in marmosets. DeepLabCut is also relatively easy to install and use, with a responsive support team from the developers (Mathis *et al.*, 2018; Lauer *et al.*, 2022). Following the successful implementation of DeepLabCut, we tried various calibration methods to obtain 3D trajectories but faced different challenges with each software, ranging from issues with corner detection and interference of the cage grid to constraints on the camera angle. We decided to use MATLAB StereoCamera

Calibrator since it worked very well with our camera configuration, and Pose3D provided the code to semi-automatically process DLC detections and use calibration parameters to triangulate detections from two cameras and obtain a 3D coordinate (Sheshadri *et al.*, 2020).

We used multi-animal DeepLabCut projects to analyse videos of a marmoset performing a sliding board task and combined two sets of body part detections from different angles to get a 3D reconstruction of the trajectory of the marmoset through camera calibration. This pipeline works very well for a single marmoset in the frame with less than 10 pixels error in detection (Figure 20) and less than 2 pixels error in 3D reconstruction. However, given that the marmosets were in different 3D reconstructed spaces, we had to devise an algorithm to align their coordinates to the real-world axes before analysing them for gaze direction or task kinematics. It's important to keep in mind that uncertainty in final estimation compounds on to previous errors from DeepLabCut detection, 3D calibration error, and aligning transformation of the 3D coordinates. Another source of error is the misalignment in recordings of paired cameras by 100 - 900 ms, which is variable over time. This is likely due to the drift in the internal clocks of the Raspberry Pis and it's a systematic and unavoidable error that might affect the 3D reconstruction to some extent based on the movement of the marmoset during segments of maximum drift.

We also trained the maDLC projects on frames from pre-task and post-task videos, and while the detection error of the body parts is quite low, the network is less accurate with assigning ID to the individuals and switches the ID when an individual moves in and out of the field of view. This is a challenge since the marmosets are not very distinctive except for the shaved segments on their tail, which indicates their within-group ID. We're trying to improve the maDLC performance by training the network on a much bigger dataset and modifying a parameter such that more body part pairwise connections are used when grouping points together.

After successfully implementing maDLC, Pose3D and 3D transformation algorithms on marmoset task videos, we wanted to analyse the gaze direction of coordinating marmoset dyads to study if they use mutual gazing and gaze following to coordinate with each other. First, we plotted the distribution of angle between head vectors of the individuals and saw that the distribution peaks at about 325° , which means that individuals most often look in the same direction by a difference of 35° (Figure 24). This distribution curve is very similar across different task conditions despite the low prosociality rate. This observation is in line with our predictions – marmosets spend most of the time looking at the experimenter who is providing them with food rewards, and hence, they end up looking in the same direction most of the time, irrespective of task condition.

We also calculated the fraction of time for which individuals' gaze intersects their own food board and that of the other individual. In line with our predictions, we see an increasing trend in the fraction of time spent looking at their own food board from prosocial to individual task, while the joint reward task lies in between (Figure 26a). We also see that marmosets spend a larger fraction of time looking at the food board of the other individual during the prosocial task than the individual task (Figure 26b). In order to check if this increase stems from attempting to coordinate and establish joint attention with the other individual or from simply following the food reward, we differentiated the gaze direction during prosocial trials based on the status of the individual – either receiver or puller. If the dyad was attempting to coordinate during the task, both should be looking at each other and by proxy, at the other's board for the same proportion of time. However, if they are just following the food reward, the puller would be looking more at the other

individual's board than the receiver. This is indeed what we see – that the individuals are mostly following the food reward (Figure 27a).

We also observe that individuals spend more time looking at the task board during successful prosocial trials, irrespective of their individual status (Figure 27b). The increase in total time spent looking at the task board can be considered as a proxy for how long they are waiting on the platforms during the trial, motivated to coordinate during the task. We can see that they spend a larger fraction of time at the task when there is a successful prosocial trial. This hints to the possibility that their low prosocial rate in the task could be because they are not able to focus on the relevant aspects of the task, similar to what was observed in preschool children who failed at the prosocial game due to high attentional demands rather than a lack of prosociality (Burkart and Rueth, 2013).

We wanted to study the kinematics of task execution as the marmosets are pulling the sliding board in the individual task condition. We also wanted to check if individuals get more efficient over the course of six trials in a session or develop a preference for the use of one hand over the other. However, we don't observe any trend in the handedness or time taken for each trial across trials within a session or across sessions (Figure 28).

We also examined whether individuals within the dyad show greater similarity in kinematics as compared to individuals outside the group. First, we compared the DTW distance measure (which is the inverse of the similarity of two time series) of the task execution within the pulls of an individual and between the pulls of a dyad in the same session. As expected, we see that the patterns of task execution are more similar within individuals in comparison to the pull across dyads (Figure 29a). Next, we compared the DTW distance metric between the pulls of real dyads versus that of pseudo dyads, and we didn't see any difference (Figure 29b). This could be because the individual task is a very simple one that can be achieved by simply gripping the handle and pulling the board close. This doesn't allow for a lot of variation in how the task can be executed. There are also additional spatial constraints on the marmosets, given that they must sit on a platform and reach out to the handle through a grid. Perhaps we would see greater similarity in dyads in a more complex task with opportunities for different strategies to succeed.

The implementation of this pipeline opens a lot of new avenues to answer different questions about marmoset behaviour without manual behavioural coding. Following this, the main goal of the project is to study behavioural synchrony in freely interacting marmosets in the pre-task and post-task videos. Once the ID switching issue is resolved, we will overlay the poses of individuals along a central axis and quantify synchrony by comparing the 3D time series of their corresponding body parts using recurrence analysis. We would expect a greater increase in behavioural synchrony and proximity after a successful prosocial task than the joint reward task, which would in turn be greater than the increase in synchrony after the individual task. We would also expect that the increase in synchrony is positively correlated with the level of prosociality in the prosocial sessions. In the longer run, this automated tracking and 3D reconstruction can be implemented in a larger enclosure to track entire groups of marmosets, study group interactions and develop behaviour classification algorithms from the pose of marmosets.

Bibliography

1. Antón, SC, Potts, R, and Aiello, LC (2014). Evolution of early Homo: An integrated biological perspective. *Science* 345, 1236828.
2. Babel, M (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J Phon* 40, 177–189.
3. Bernieri, FJ (1988). Coordinated movement and rapport in teacher-student interactions. *J Nonverbal Behav* 12, 120–138.
4. Bernieri, FJ, and Rosenthal, R (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In: *Fundamentals of Nonverbal Behavior*, New York, NY, US: Cambridge University Press, 401–432.
5. Bluhm, E, and Hedrick, T *EasyWand Camera Calibration User Guide*.
6. Boker, SM, Rotondo, JL, Xu, M, and King, K (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychol Methods* 7, 338–355.
7. Bouget, J-Y (2022). *Camera Calibration Toolbox for Matlab*, CaltechDATA.
8. Brennan, SE, and Clark, HH (1996). Conceptual pacts and lexical choice in conversation. *J Exp Psychol Learn Mem Cogn* 22, 1482–1493.
9. Brügger, RK, Willems, EP, and Burkart, JM (2022). Looking out for each other: coordination and turn taking in common marmoset vigilance. *Anim Behav*.
10. Burkart, J, and Heschl, A (2006). Geometrical gaze following in common marmosets (*Callithrix jacchus*). *J Comp Psychol* 120, 120–130.
11. Burkart, JM, Adriaense, JEC, Brügger, RK, Miss, FM, Wierucka, K, and van Schaik, CP (2022). A convergent interaction engine: vocal communication among marmoset monkeys. *Philos Trans R Soc B Biol Sci* 377, 20210098.
12. Burkart, JM, and Finkenwirth, C (2015). Marmosets as model species in neuroscience and evolutionary anthropology. *Neurosci Res* 93, 8–19.
13. Burkart, JM, Hrdy, SB, and Van Schaik, CP (2009). Cooperative breeding and human cognitive evolution. *Evol Anthropol Issues News Rev* 18, 175–186.
14. Burkart, JM, and Rueth, K (2013). Preschool Children Fail Primate Prosocial Game Because of Attentional Task Demands. *PLOS ONE* 8, e68440.
15. Burkart, JM, and van Schaik, CP (2020). Marmoset prosociality is intentional. *Anim Cogn* 23, 581–594.
16. Cao, Z, Simon, T, Wei, S-E, and Sheikh, Y (2017). Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. 7291–7299.
17. Chartrand, TL, and Bargh, JA (1999). The chameleon effect: The perception–behavior link and social interaction. *J Pers Soc Psychol* 76, 893–910.
18. Coco, MI, and Dale, R (2014). Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Front Psychol* 5.
19. Creaven, A-M, Skowron, EA, Hughes, BM, Howard, S, and Loken, E (2014). Dyadic concordance in mother and preschooler resting cardiovascular function varies by risk status. *Dev Psychobiol* 56, 142–152.
20. Daoudi-Simison, S, O’Sullivan, E, Moat, G, Lee, PC, and Buchanan-Smith, HM (2023). Do mixed-species groups of capuchin (*Sapajus apella*) and squirrel monkeys (*Saimiri sciureus*) synchronize their behaviour? *Philos Trans R Soc B Biol Sci* 378, 20220111.
21. Delaherche, E, Chetouani, M, Mahdhaoui, A, Saint-Georges, C, Viaux, S, and Cohen, D (2012). Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines. *IEEE Trans Affect Comput* 3, 349–365.
22. Durantón, C, and Gaunet, F (2016). Behavioural synchronization from an ethological perspective: Overview of its adaptive value. *Adapt Behav* 24, 181–191.
23. Ebrahimi, AS, Orłowska-Feuer, P, Huang, Q, Zippo, AG, Martial, FP, Petersen, RS, and Storchi, R (2023). Three-dimensional unsupervised probabilistic pose reconstruction (3D-UPPER) for freely moving animals. *Sci Rep* 13, 155.
24. Emery, NJ (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci Biobehav Rev* 24, 581–604.
25. Feldman, R (2012). Bio-behavioral Synchrony: A Model for Integrating Biological and Microsocial Behavioral Processes in the Study of Parenting. *Parenting* 12, 154–164.
26. Ferrer, E, and Helm, JL (2013). Dynamical systems modeling of physiological coregulation in dyadic interactions. *Int J Psychophysiol* 88, 296–308.
27. Finkenwirth, C, and Burkart, JM (2017). Long-term-stability of relationship structure in family groups of

- common marmosets, and its link to proactive prosociality. *Physiol Behav* 173, 79–86.
28. Finkenwirth, C, and Burkart, JM (2018). Why help? Relationship quality, not strategic grooming predicts infant-care in group-living marmosets. *Physiol Behav* 193, 108–116.
 29. Finkenwirth, C, Martins, E, Deschner, T, and Burkart, JM (2016). Oxytocin is associated with infant-care behavior and motivation in cooperatively breeding marmoset monkeys. *Horm Behav* 80, 10–18.
 30. Fuhrmann, D, Ravignani, A, Marshall-Pescini, S, and Whiten, A (2014). Synchrony and motor mimicking in chimpanzee observational learning. *Sci Rep* 4, 5283.
 31. Gaziv, G, Noy, L, Liron, Y, and Alon, U (2017). A reduced-dimensionality approach to uncovering dyadic modes of body motion in conversations. *PLOS ONE* 12, e0170786.
 32. Gosztolai, A, and Ramdya, P (2022). Connecting the dots in ethology: applying network theory to understand neural and animal collectives. *Curr Opin Neurobiol* 73, 102532.
 33. Graving, JM, Chae, D, Naik, H, Li, L, Koger, B, Costelloe, BR, and Couzin, ID (2019). DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* 8, e47994.
 34. Guerreiro Martins, EM, Moura, AC de A, Finkenwirth, C, Griesser, M, and Burkart, JM (2019). Food sharing patterns in three species of callitrichid monkeys (*Callithrix jacchus*, *Leontopithecus chrysomelas*, *Saguinus midas*): Individual and species differences. *J Comp Psychol* 133, 474–487.
 35. Guitton, D, and Volle, M (1987). Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *J Neurophysiol* 58, 427–459.
 36. Gvirts, HZ, and Perlmutter, R (2020). What Guides Us to Neurally and Behaviorally Align With Anyone Specific? A Neurobiological Model Based on fNIRS Hyperscanning Studies. *The Neuroscientist* 26, 108–116.
 37. Healey, PGT, Purver, M, and Howes, C (2014). Divergence in Dialogue. *PLOS ONE* 9, e98598.
 38. Hedrick, TL (2008). Software techniques for two- and three-dimensional kinematic measurements of biological and biomimetic systems. *Bioinspir Biomim* 3, 034001.
 39. Herrmann, E, Hernández-Lloreda, MV, Call, J, Hare, B, and Tomasello, M (2010). The Structure of Individual Differences in the Cognitive Abilities of Children and Chimpanzees. *Psychol Sci* 21, 102–110.
 40. Hove, MJ (2008). Shared circuits, shared time, and interpersonal synchrony. *Behav Brain Sci* 31, 29–30.
 41. Hove, MJ, and Risen, JL (2009). It's All in the Timing: Interpersonal Synchrony Increases Affiliation. *Soc Cogn* 27, 949–960.
 42. Jackson, BE, Evangelista, DJ, Ray, DD, and Hedrick, TL (2016). 3D for the people: multi-camera motion capture in the field with consumer-grade cameras and open source software. *Biol Open* 5, 1334–1342.
 43. Karashchuk, P, Rupp, KL, Dickinson, ES, Walling-Bell, S, Sanders, E, Azim, E, Brunton, BW, and Tuthill, JC (2021). Anipose: A toolkit for robust markerless 3D pose estimation. *Cell Rep* 36, 109730.
 44. King, AJ, and Cowlshaw, G (2009). All together now: behavioural synchrony in baboons. *Anim Behav* 78, 1381–1387.
 45. Koski, SE, and Burkart, JM (2015). Common marmosets show social plasticity and group-level similarity in personality. *Sci Rep* 5, 8878.
 46. Lang, M, Bahna, V, Shaver, JH, Reddish, P, and Xygalatas, D (2017). Sync to link: Endorphin-mediated synchrony effects on cooperation. *Biol Psychol* 127, 191–197.
 47. Lauer, J, Zhou, M, Ye, S, Menegas, W, Schneider, S, Nath, T, Rahman, MM, Di Santo, V, Soberanes, D, Feng, G, *et al.* (2022). Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nat Methods* 19, 496–504.
 48. Luxem, K, Sun, JJ, Bradley, SP, Krishnan, K, Yttri, E, Zimmermann, J, Pereira, TD, and Laubach, M (2023). Open-source tools for behavioral video analysis: Setup, methods, and best practices. *eLife* 12, e79305.
 49. Macrae, CN, Duffy, OK, Miles, LK, and Lawrence, J (2008). A case of hand waving: Action synchrony and person perception. *Cognition* 109, 152–156.
 50. Marks, M, Jin, Q, Sturman, O, von Ziegler, L, Kollmorgen, S, von der Behrens, W, Mante, V, Bohacek, J, and Yanik, MF (2022). Deep-learning-based identification, tracking, pose estimation and behaviour classification of interacting primates and mice in complex environments. *Nat Mach Intell* 4, 331–340.
 51. Marwan, N, Carmen Romano, M, Thiel, M, and Kurths, J (2007). Recurrence plots for the analysis of complex systems. *Phys Rep* 438, 237–329.
 52. Massen, JJM, Šlipogor, V, and Gallup, AC (2016). An Observational Investigation of Behavioral Contagion in Common Marmosets (*Callithrix jacchus*): Indications for Contagious Scent-Marking. *Front Psychol* 7.
 53. Mathis, A, Mamidanna, P, Cury, KM, Abe, T, Murthy, VN, Mathis, MW, and Bethge, M (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* 21, 1281–1289.
 54. Meert, W, Hendrickx, K, Van Craendonck, T, Robberechts, P, Blockeel, H, and Davis, J (2020). DTAIDistance, Zenodo.
 55. Miss, FM, and Burkart, JM (2018). Corepresentation During Joint Action in Marmoset Monkeys (*Callithrix*

- jacchus). *Psychol Sci* 29, 984–995.
56. Mitchell, JF, Reynolds, JH, and Miller, CT (2014). Active Vision in Marmosets: A Model System for Visual Neuroscience. *J Neurosci* 34, 1183–1194.
 57. Mogan, R, Fischer, R, and Bulbulia, JA (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *J Exp Soc Psychol* 72, 13–20.
 58. Moro, M, Marchesi, G, Hesse, F, Odone, F, and Casadio, M (2022). Markerless vs. Marker-Based Gait Analysis: A Proof of Concept Study. *Sensors* 22, 2011.
 59. Müller, M (2007). Dynamic Time Warping. In: *Information Retrieval for Music and Motion*, Berlin, Heidelberg: Springer, 69–84.
 60. Nath, T, Mathis, A, Chen, AC, Patel, A, Bethge, M, and Mathis, MW (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc* 14, 2152–2176.
 61. Nishikawa, M, Suzuki, M, and Sprague, DS (2021). Activity synchrony and travel direction synchrony in wild female Japanese macaques. *Behav Processes* 191, 104473.
 62. de Oliveira Terceiro, FE, Willems, EP, Araújo, A, and Burkart, JM (2021). Monkey see, monkey feel? Marmoset reactions towards conspecifics' arousal. *R Soc Open Sci* 8, 211255.
 63. Olsen, NL, Markussen, B, and Raket, LL (2018). Simultaneous inference for misaligned multivariate functional data. *J R Stat Soc Ser C Appl Stat* 67, 1147–1176.
 64. Ota, N, Gahr, M, and Soma, M (2015). Tap dancing birds: the multimodal mutual courtship display of males and females in a socially monogamous songbird. *Sci Rep* 5, 16614.
 65. Palumbo, RV, Marraccini, ME, Weyandt, LL, Wilder-Smith, O, McGee, HA, Liu, S, and Goodwin, MS (2017). Interpersonal Autonomic Physiology: A Systematic Review of the Literature. *Personal Soc Psychol Rev* 21, 99–141.
 66. Pandey, S, Simhadri, S, and Zhou, Y (2020). Rapid Head Movements in Common Marmoset Monkeys. *iScience* 23, 100837.
 67. Pereira, TD, Tabris, N, Matsliah, A, Turner, DM, Li, J, Ravindranath, S, Papadoyannis, ES, Normand, E, Deutsch, DS, Wang, ZY, *et al.* (2022). SLEAP: A deep learning system for multi-animal pose tracking. *Nat Methods* 19, 486–495.
 68. Perrett, DI, and Mistlin, AJ (1990). Perception of facial characteristics by monkeys. In: *Comparative Perception, Vol. 2: Complex Signals*, Oxford, England: John Wiley & Sons, 187–215.
 69. Phaniraj, N, Brügger, RK, and Burkart, JM (2023). Marmosets mutually compensate for differences in rhythms when coordinating vigilance. 2023.09.28.559895.
 70. Ramseyer, F, and Tschacher, W (2010). Nonverbal Synchrony or Random Coincidence? How to Tell the Difference. In: *Development of Multimodal Interfaces: Active Listening and Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23–27, 2009, Revised Selected Papers*, ed. A Esposito, N Campbell, C Vogel, A Hussain, and A Nijholt, Berlin, Heidelberg: Springer, 182–196.
 71. Reddish, P, Fischer, R, and Bulbulia, J (2013). Let's Dance Together: Synchrony, Shared Intentionality and Cooperation. *PLOS ONE* 8, e71182.
 72. Rennung, M, and Göritz, AS (2016). Prosocial Consequences of Interpersonal Synchrony. *Z Für Psychol* 224, 168–189.
 73. Richardson, JRLG, Robert W Isenhower, Kerry L Marsh, RC Schmidt, Michael J (2005). The Interpersonal Phase Entrainment of Rocking Chair Movements. In: *Studies in Perception and Action VIII*, Psychology Press.
 74. Ruch, H, Zürcher, Y, and Burkart, JM (2018). The function and mechanism of vocal accommodation in humans and other primates. *Biol Rev* 93, 996–1013.
 75. Schmidt, RC, Morr, S, Fitzpatrick, P, and Richardson, MJ (2012). Measuring the Dynamics of Interactional Synchrony. *J Nonverbal Behav* 36, 263–279.
 76. Schmidt, RC, and Turvey, MT (1994). Phase-entrainment dynamics of visually coupled rhythmic movements. *Biol Cybern* 70, 369–376.
 77. Schoenherr, D, Paulick, J, Worrack, S, Strauss, BM, Rubel, JA, Schwartz, B, Deisenhofer, A-K, Lutz, W, Stangier, U, and Altmann, U (2019). Quantification of nonverbal synchrony using linear time series analysis methods: Lack of convergent validity and evidence for facets of synchrony. *Behav Res Methods* 51, 361–383.
 78. Shepherd, S (2010). Following Gaze: Gaze-Following Behavior as a Window into Social Cognition. *Front Integr Neurosci* 4.
 79. Sheshadri, S, Dann, B, Hueser, T, and Scherberger, H (2020). 3D reconstruction toolbox for behavior tracked with multiple cameras. *J Open Source Softw* 5, 1849.
 80. Shockley, K, Baker, AA, Richardson, MJ, and Fowler, CA (2007). Articulatory constraints on interpersonal postural coordination. *J Exp Psychol Hum Percept Perform* 33, 201–208.
 81. Shockley, K, Santana, M-V, and Fowler, CA (2003). Mutual interpersonal postural constraints are involved in

- cooperative conversation. *J Exp Psychol Hum Percept Perform* 29, 326–332.
82. Snowdon, C, and Elowson, AM (2001). “BABBLING” IN PYGMY MARMOSETS: DEVELOPMENT AFTER INFANCY. *Behaviour* 138, 1235–1248.
 83. Snowdon, CT (2001). Social processes in communication and cognition in callitrichid monkeys: a review. *Anim Cogn* 4, 247–257.
 84. Spadacenta, S, Dicke, PW, and Thier, P (2019). Reflexive gaze following in common marmoset monkeys. *Sci Rep* 9, 15292.
 85. Stephens, GJ, Silbert, LJ, and Hasson, U (2010). Speaker–listener neural coupling underlies successful communication. *Proc Natl Acad Sci* 107, 14425–14430.
 86. Stratford, T, Lal, S, and Meara, A (2012). Neuroanalysis of Therapeutic Alliance in the Symptomatically Anxious: The Physiological Connection Revealed between Therapist and Client. *Am J Psychother* 66, 1–21.
 87. Sun, JJ, Karashchuk, L, Dravid, A, Ryou, S, Fereidooni, S, Tuthill, JC, Katsaggelos, A, Brunton, BW, Gkioxari, G, Kennedy, A, *et al.* (2023). BKinD-3D: Self-Supervised 3D Keypoint Discovery From Multi-View Videos.
 88. Takahashi, DY, Narayanan, DZ, and Ghazanfar, AA (2013). Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys. *Curr Biol* 23, 2162–2168.
 89. Tomasello, M, Carpenter, M, Call, J, Behne, T, and Moll, H (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behav Brain Sci* 28, 675–691.
 90. Tschacher, W, Rees, GM, and Ramseyer, F (2014). Nonverbal synchrony and affect in dyadic interactions. *Front Psychol* 5.
 91. Tunçgenç, B, and Cohen, E (2016). Movement Synchrony Forges Social Bonds across Group Divides. *Front Psychol* 7.
 92. Valdesolo, P, Ouyang, J, and DeSteno, D (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *J Exp Soc Psychol* 46, 693–695.
 93. Vargas-Irwin, CE, Shakhnarovich, G, Yadollahpour, P, Mislow, JMK, Black, MJ, and Donoghue, JP (2010). Decoding Complete Reach and Grasp Actions from Local Primary Motor Cortex Populations. *J Neurosci* 30, 9659–9669.
 94. Voelkl, B, and Huber, L (2007). Imitation as Faithful Copying of a Novel Technique in Marmoset Monkeys. *PLOS ONE* 2, e611.
 95. Wallot, S, Roepstorff, A, and Mønster, D (2016). Multidimensional Recurrence Quantification Analysis (MdrQA) for the Analysis of Multidimensional Time-Series: A Software Implementation in MATLAB and Its Application to Group-Level Data in Joint Action. *Front Psychol* 7.
 96. Walter, T, and Couzin, ID (2021). TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. *eLife* 10, e64000.
 97. Wheatley, T, Kang, O, Parkinson, C, and Looser, CE (2012). From Mind Perception to Mental Connection: Synchrony as a Mechanism for Social Understanding. *Soc Personal Psychol Compass* 6, 589–606.
 98. Xing, F, Sheffield, AG, Jadi, MP, Chang, SWC, and Nandy, AS (2024). Automated 3D analysis of social head-gaze behaviors in freely moving marmosets. 2024.02.16.580693.
 99. Zhang, Z (2000). A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22, 1330–1334.
 100. van der Zouwen, CI, Boutin, J, Fougère, M, Flaive, A, Vivancos, M, Santuz, A, Akay, T, Sarret, P, and Ryczko, D (2021). Freely Behaving Mice Can Brake and Turn During Optogenetic Stimulation of the Mesencephalic Locomotor Region. *Front Neural Circuits* 15.
 101. Zürcher, Y, and Burkart, JM (2017). Evidence for Dialects in Three Captive Populations of Common Marmosets (*Callithrix jacchus*). *Int J Primatol* 38, 780–793.

Appendix

The following tables contain the estimated parameters of camera calibration for the two pairs of cameras from MATLAB StereoCameraCalibration. The camera intrinsic and extrinsic parameters were estimated by detecting corners in paired checkerboard images. The mean reprojection error measured in pixels is a measure of the accuracy of the estimation.

Camera model

Appendix 1. Camera model features estimated for each pair of cameras

Features	Left Cameras	Right Cameras
No. of distortion coefficients	3	3
Compute skew	1	1
Compute tangential distortion	1	1
No. of boards	241	238
Checkerboard corners	28	28
World units	mm	mm
Mean Reprojection Error	1.4716	0.6998

Camera intrinsics

Appendix 2. Camera intrinsics estimated for each pair of cameras

Parameters	Left Cameras	Right Cameras
Camera 1 Intrinsics		
Focal length	[2711.6 2668.7]	[2609.9 2471.6]
Principal Point	[-68.165 158.146]	[799.908 420.996]
Image size	[1080 960]	[1080 960]
Radial distortion	[-0.6341 1.4237 -3.2059]	[-0.5655 0.3410 0.0605]
Tangential distortion	[0.011 -0.0016]	[0.0055 -0.0028]
Skew	19.5268	-18.9653
K	$\begin{bmatrix} 2711.6 & 19.52 & -68.1 \\ 0 & 2668.7 & 158.1 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2609.9 & -18.96 & 799.9 \\ 0 & 2471.6 & 420.9 \\ 0 & 0 & 1 \end{bmatrix}$
Mean Reprojection Error	1.3068	0.6657
Camera 2 Intrinsics		
Focal length	[3056.7 2412.1]	[2809.8 2448.5]
Principal Point	[-718.479 307.2416]	[1163.2 364.079]
Image size	[1080 960]	[1080 960]

Radial distortion	$[-0.5153 \ 0.3291 \ -0.3197]$	$[-0.5907 \ 0.7318 \ -1.3541]$
Tangential distortion	$[0.0018 \ 0.0274]$	$[0.0012 \ -0.0284]$
Skew	83.7254	-27.3469
K	$\begin{bmatrix} 3056.7 & 83.72 & -718.49 \\ 0 & 2412.1 & 307.24 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 2809.8 & -27.34 & 1163.3 \\ 0 & 2448.5 & 364.07 \\ 0 & 0 & 1 \end{bmatrix}$
Mean Reprojection Error	1.6365	0.7339

Inter-camera geometry

Appendix 3. Geometric relation of Camera 2 with respect to Camera 1

Parameters	Left Cameras	Right Cameras
Translation matrix	$[1288.5 \ -15.879 \ 148.983]$	$[-1.143 \ -119.637 \ 288.81]$
Rotation matrix	$\begin{bmatrix} 0.0839 & -0.0321 & -0.5421 \\ 0.0013 & 0.9984 & -0.0571 \\ 0.5431 & 0.0472 & 0.8384 \end{bmatrix}$	$\begin{bmatrix} 0.8115 & -0.2139 & 0.5438 \\ 0.1884 & 0.9767 & 0.1030 \\ -0.5531 & 0.0189 & 0.8329 \end{bmatrix}$
A	$\begin{bmatrix} 0.0839 & -0.0321 & -0.5421 & 1288.5 \\ 0.0013 & 0.9984 & -0.0571 & -15.879 \\ 0.5431 & 0.0472 & 0.8384 & 148.983 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.8115 & -0.2139 & 0.5438 & -1.1434 \\ 0.1884 & 0.9767 & 0.1030 & -119.637 \\ -0.5531 & 0.0189 & 0.8329 & 288.81 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Fundamental matrix	$\begin{bmatrix} -0.000 & -0.000 & 0.0013 \\ -0.0001 & -0.000 & -0.4858 \\ 0.0317 & 0.4714 & -7.7859 \end{bmatrix}$	$\begin{bmatrix} 0.000 & -0.000 & -0.0301 \\ -0.0001 & -0.000 & 0.5041 \\ -0.0238 & -0.4121 & 28.678 \end{bmatrix}$
Essential matrix	$\begin{bmatrix} -8.8 & -149.5 & -4.8 \\ -574.7 & -65.6 & -1161.0 \\ 15.0 & 1285.9 & -82.1 \end{bmatrix}$	$\begin{bmatrix} 11.8 & -284.3 & -129.4 \\ -398.1 & -40.1 & 1109.3 \\ -118.4 & -1142.3 & -52.7 \end{bmatrix}$