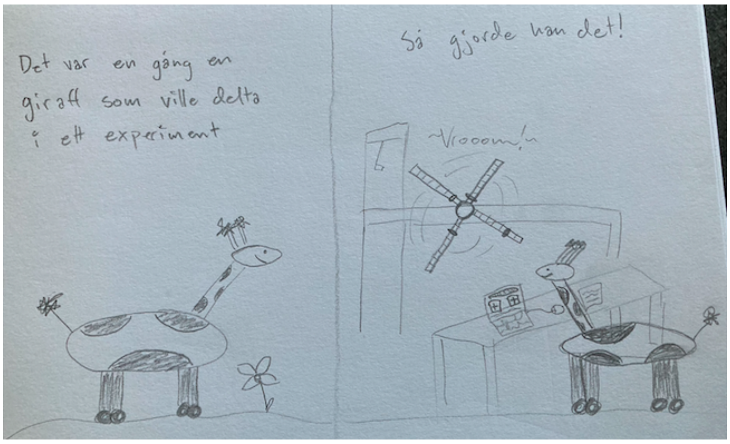
UNIVERSITY OF GOTHENBURG

DEPARTMENT OF APPLIED IT

OPTIMISATION AND UNDERSTANDING OF TECHNOLOGICAL ARTEFACTS

An Experimental Study Investigating the Role of Causal Cognition and Social Learning in Micro-society Paradigms



Albin Högberg

Eva Iliefski-Janols

Thesis:	15 hp
Program:	Cognitive science
Level:	First Cycle
Year:	2022
Supervisor:	Gustaf Lindblad
Examiner:	Pierre Gander
Report nr:	2022:090

Abstract

Cumulative technological culture - how humans can build upon and improve the technology and tools of earlier generations - has been regarded as a main contributing factor for *Homo sapiens'* success as a species. In this study we investigate the role of causal cognition and social learning in the optimisation of technological artefacts. We do so through an experimental study where participants optimise a wheel system. Our results show that participants' causal understanding of the wheel increased as a consequence of the optimisation of the wheel system. The study is a partial replication of a former study by Derex et al. (2019) with further development in the experimental design. The implications of our findings questions the validity of the former experiment by Derex et al. and gives further support in investigations of the role of social learning and causal cognition, and the relevance of micro-society paradigms in research concerning cumulative technological culture.

Keywords

Cumulative Technological Culture, Cognitive Niche Hypothesis, Cultural Niche Hypothesis, Social Learning, Causal Cognition, Micro-society Paradigms

Optimering och Förståelse av Teknologiska Artefakter: En Experimentell Studie som Undersöker Rollen av Kausal Kognition och Socialt Lärande i Mikrosamhällesparadigm

Sammanfattning

Kumulativ teknologisk kultur - hur människor bygger på och förbättrar teknik och verktyg från tidigare generationer - har betraktats som en huvudsaklig bidragande faktor för Homo sapiens framgång som art. I denna studie undersöker vi rollen av kausal kognition och socialt lärande i optimeringen av tekniska artefakter. Detta gör vi genom en experimentell studie där deltagarna optimerar ett hjulsystem. Våra resultat visar att deltagarnas kausala förståelse för hjulet ökade som en konsekvens av optimering av hjulsystemet. Studien är delvis en replikering av en tidigare studie, gjord av Derex et al. (2019), men även en vidareutveckling av experimentdesignen. Implikationerna av våra resultat ifrågasätter giltigheten av slutsatserna som baserats på experimentet av Derex et al. och ger samtidigt stöd i vidare undersökning av rollen av socialt lärande och kausal kognition samt relevansen av mikrosamhällesparadigm i forskning om kumulativ teknologisk kultur.

Nyckelord

Kumulativ Teknologisk Kultur, Kognitiva Nische Hypotesen, Kulturella Nische Hypotesen, Socialt Lärnande, Kausal Kognition, Mikrosamhällesparadigm

Foreword

This project was started in January 2020 by Albin Högberg, in collaboration with Anders Högberg and Gustaf Lindblad, but was put to a halt by the COVID-19 pandemic. In May 2021 Eva Iliefski-Janols joined the project and the planning started. From there both authors have been equally involved in the planning and execution of the experimental study. In the writing process, Albin Högberg had a larger contribution to the method and the discussion chapters. Eva Iliefski-Janols had a larger contribution to the theory and the results chapters. All other chapters were written through an equal contribution.

Thanks

Many are to be credited in the development of this study. First, we would like to thank Alexander Wiig, Sebastian Andreasson and Mattias Rost for the building and development of the apparatus and the acquisition of the materials used in the study. Secondly, we would like to thank Alexander Almér, Anders Högberg, Marlize Lombard, Peter Gärdenfors and William Hedley Thompson for all of our rewarding discussions. Thirdly, we extend a thanks to Jenelyn Aggerstam and Karin Pettersson for their wonderful administrative help. We would also like to thank all of the participants who took part in the study, who made this all possible.

A special thanks goes out to Gustaf Lindblad for his guidance and advice in his role as our supervisor. Thank you for all the wonderful tips on music, for letting us meet your daughter and pet your cat.

Table of contents

1. Introduction

Homo sapiens successfully inhabits the largest and most diverse range of habitats across the world, despite being biologically ill suited for many of them (Boyd & Richerson, 1985). We have built weapons for hunting in absence of sharp teeth or claws. We have built houses to shelter us from harsh weather and dangers, to combat our lack of thick fur or hard skin. It is the different tools we have created that have allowed humankind to thrive in a vast range of environments. Tools are constantly being modified, further developed and improved to better fit the users' desires. This is referred to as cumulative technological culture, to build upon and improve the technology of earlier generations. Inter and intra generational teaching, learning through each other and developing what others before had already created, allows technology to progress further than what any single individual could create alone (Boyd & Richerson, 1998; Boyd et al., 2011; Richerson & Boyd, 2005; Tomasello, 1999; Tomasello et al., 1993).

How can we explain the complexity of technology and cumulative culture? A causal explanation must be able to handle the complexities of human culture - it is cumulative, socially transferred within and between populations, across generations and varies both within and between populations (Legare, 2017). Creating and using tools to solve problems requires a degree of causal understanding of the world (Bender & Beller, 2019; De Oliveira et al., 2019; Gärdenfors & Lombard, 2020; Tooby & De Vore, 1987). Ensuring that tools and necessary knowledge about them persist in populations requires some teaching and learning between members of a group (Boyd & Richerson, 1998; Boyd et al. 2011; Tomasello et al., 2005). Some tools need cooperation between members to be used efficiently and cooperation in turn requires certain socio-cognitive abilities, such as theory of mind and shared intentionality (Moll et al., 2020; Tomasello et al., 2005; Whiten & Erdal, 2012). In modern society much of our behaviour and our inventions are taken for granted and deemed natural. In reality, they should still be regarded as impressive feats and decidedly unusual as no other species have developed tools quite as complex as ours. Should all technology and corresponding knowledge magically disappear, many of the environments presently inhabited by humans would become inhospitable. Since *Homo sapiens* as a species are obligatory tool users, no human population would survive without the usage of tools (Shea, 2017).

In the study “Causal understanding is not necessary for the improvement of culturally evolving technology” Derex et al. (2019) aimed to test the cognitive niche hypothesis and to see whether causal understanding is necessary for improving tools and technology. The study consisted of an experiment with micro-society paradigms where participants’ task was to improve a physical wheel system of two dimensions across artificial generations. Derex et al. proposes in their conclusions that complex technologies need not result from a development of understanding of the technology at hand. Rather, they conclude this development to take place within the boundaries of learning through trial and error. Further, they argue that explicit tool usage and technological advancements can emerge solely from the accumulation of small improvements made in the course of generational overlaps.

Osiurak et al. (2021) replicated the experiment by Derex et al. (2019) and showed that the conclusions made by Derex et al. were not made from a sufficient experimental design. In their replication, Osiurak et al. mainly focused on implementation of gained knowledge and developed the experimental design with regards to the transfer of knowledge to a similar situation. They concluded that if the experimental conditions were altered with regards to implementation of knowledge, the conclusions made by Derex et al. (2019) were insufficient.

1.1 Purpose and hypotheses

From the experimental study by Derex et al. (2019) we identified weaknesses in the experimental design with regards to social learning, causal cognition and cultural transmittance. This study partly serves as a replication of the experiment conducted by Derex et al., with further grounds in the replication by Osiurak et al. (2021), and aims to be a methodological critique regarding how micro-society paradigms are implemented with lack of natural circumstances of social learning. Furthermore, it serves as a development of the experimental design by adding aspects of social learning and cultural transmittance. By adding two new variables to the experimental design, split over two separate conditions, this study further investigates the role of social learning in micro-society paradigms in cumulative technological culture research. The variables were added in accordance with previous research on social learning and optimisation of artefacts and tools (Caldwell & Millen, 2008; Poulin-Dubois & Brosseau-Liard, 2016; Sabbagh & Koenig, 2013; Stenberg, 2009; Stenberg & Hagekull, 2007; Thompson et al., 2022; Tomasello, 2016). The purpose of this study is to examine whether the results from the study by Derex et al. (2019) are plausible and to further investigate the relevance of conclusions regarding cumulative culture based on micro-society experiments. Additionally, this study allows us to further develop the research on cumulative culture and culturally improving artefacts.

Two hypotheses were created in accordance with the purpose of the study:

1. Optimisation of a physical artefact will be followed by an increase of knowledge regarding the technological causality of the artefact.
2. The introduction of aspects of a social learning environment by prompting vocal verbalisation and a change in transmissions of knowledge between in-group participants will affect participants' results.

2. Theory

2.1 Cumulative Technological Culture

The term cumulative technological culture (henceforth referred to as CTC) was coined by Boyd and Richersson (1985). Cumulative culture is highly varied across cultures of the world. Adaptations that support cumulative culture are hypothesised to be a universal feature of human psychology, yet flexible enough to support the variations of humans capacity for acquiring different skill sets and behaviours (Legare, 2017). The cumulative aspects of culture is a social phenomenon, requiring social interaction and exchange to be considered cultural (Laland et al., 2011; O'Brien & Bentley, 2021). Research regarding CTC is multidisciplinary, combining research from e.g. evolutionary psychology, archaeology, anthropology and cognitive science. The interdisciplinarity of the field has given rise to differing theories as to how to approach and understand the causality and features of the phenomenon. One divide is that between the cognitive niche hypothesis; which emphasise non-social and cognitive factors, such as technological and causal reasoning as the driving force behind CTC, and the cultural niche hypothesis; which emphasise social and cultural factors as the driving force behind CTC. An additional divide is between those who see these two hypotheses as incompatible and contradictory and those who consider them beneficial and complementary to each other.

To understand CTC it is important to create an explanation of the present features and the evolutionary driving forces of its emergence. Social cognitive abilities such as theory of mind (Osiurak et al., 2020; Premack & Woodruff, 1978; Whiten & Erdal, 2012), shared intentionality (Moll et al., 2020; Tomasello et al., 2005), and language (Kirby & Tamariz, 2022; Pinker, 2010; Tomasello, 2000) have been identified as important factors for the development of greater understanding and cooperation with others. Human tools are adapted to their local environment and to live and thrive in a specific local environment requires complex toolmaking and knowledge of the habitat, besides cooperation with other people (Boyd et al., 2011). Tools and technological inventions were not invented and used by one individual, but spread to wider populations where they persisted to be improved and/or modified over time and generations. This has led to prevalent theories of social learning and sharing of knowledge as important factors for cumulative culture (Boyd et al. 2011; Tomasello et al., 2005). Creation and usage of tools require sufficient cognitive abilities, hence the important role of technological understanding and ability for causal reasoning (Bender & Beller, 2019; De Oliveira et al., 2019; Gärdenfors & Lombard, 2020; Lombard & Gärdenfors, 2017; Tooby & De Vore, 1987). While this sophistication of technological understanding so far seems to be unique to humans, comparative studies with species such as New Caledonian crows (Hunt & Gray, 2003, 2004; Taylor et al., 2012) and chimpanzees (Whiten & Van Schaik, 2007; Yamamoto et al., 2013) show that some features of cumulative culture and social learning can be found in other species as well.

2.1.1 Cognitive Niche Hypothesis

The cognitive niche hypothesis proposes an explanation to behaviours that are either unique to or considerably more developed in *Homo sapiens* compared to other species. The hypothesis was first proposed by Tooby and DeVore (1987) reasoning that humans exploit a cognitive niche in order to adapt to local and differing environments. They suggested that humans have evolved a sort of “improvisational intelligence” in contrast to other animals' “dedicated intelligence”. Dedicated intelligence allows domain-specific learning and adaptations suited for particular environments. Improvisational intelligence allows humans to make inferences about the world based on intuitive understandings of e.g. physics, biology and psychology. The cognitive niche describes a modular view of the mind where improvisational inferences happen through internal mental models arrangeable in a range of combinations as-they-go. Behaviours originating from improvisational intelligence can adapt and change over individuals' lifetimes with feedback from others and from the environment with trial-and-error testing, instead of random mutations and the slower and gradual evolutionary changes that we see in other species adaptation for their survival in specific environments (Tooby & DeVore, 1987).

Steven Pinker built upon the hypothesis of how hominids evolved to specialise in the cognitive niche by studying causal structures (technical know-how), cooperation among non-kin and grammatical language (Pinker, 2010). The technical know-how is based on intuitive theories, understanding of causal relationships and psychological factors such as internal beliefs and desires. Cooperation among non-kin can be supported by how humans build and nurture relationships and cooperate with others. Grammatical language paves the way for arbitrary conceptual and mutual understandings of each other and serves an advantage for transferring knowledge and information to one another. Pinker further argues that cognition, language and sociality altered the social environments for ancestral hominids in a coevolutionary process that allowed the cognitive niche to develop. This stance, primarily developed by Pinker, argues for a more social version of the cognitive niche hypothesis than that of the original version proposed by Tooby and DeVore (1987).

Osiurak and Reynaud (2019) proposed a technological reasoning hypothesis, an off-shot from the cognitive niche hypothesis, that puts technical reasoning as the primary driving force of CTC. They argue that while CTC is a both technical and social phenomenon, there has not been enough emphasis on the technological aspect - a conclusion supported by neuroimaging and neuropsychological research (Osiurak & Reynaud, 2019). Technical reasoning, as used in this context, is not synonymous with causal reasoning or understanding but relates to a form of causal reasoning concerned with technical content (Osiurak et al. 2020). Technical reasoning is defined as the ability to reason implicitly about physical object properties. Continuing, it is both analogical (knowledge can be transferred from one situation to another) and causal (knowledge can be used to predict the effects on the environment) (Osiurak & Reynaud, 2019, 2021; Osiurak et al., 2020). Osiurak and Reynaud (2019) proposed that technical reasoning allows humans to acquire and generate a great amount of information and technical knowledge. Individuals acquire techniques over time depending on opportunities available in their environment, which in turn leads to a great individual variation. While any one individual does not have the capacity to acquire all knowledge or create all tools in a single lifetime, any individual has the potential to learn any technique. The technological reasoning hypothesis supports that CTC is specific to our species and that technical reasoning skills have been a crucial divider between our species and other animals (Osiurak & Reynaud, 2020).

2.1.2 Cultural Niche Hypothesis

The cultural niche hypothesis, sometimes referred to as the socio-cognitive niche (Whiten & Erdal, 2012), proposes a coevolutionary process of human causal reasoning abilities along with more social cognitive aspects like cooperation and culture. From the cultural niche hypothesis, it follows that causal reasoning alone is not sufficient to develop the sophisticated tools and knowledge referred to in CTC. Tools and knowledge have not only developed, but spread to wider populations to persist and to later be improved upon in a continuous cycle (Boyd & Richerson, 1985; Boyd et al., 2011; Richerson & Boyd, 2005).

While the cognitive niche hypothesis suggests that causal and technical reasoning abilities should be sufficient alone for both creation and improvement of tools, some observational accounts support the view that cultural transmissions are necessary. One of the most telling documented examples is from an isolated group of Polar Inuit of Northwest Greenland. In 1853 explorers Elisha Kane and Isaac Hayes wintered there and reported that the group lacked kayaks, bows and arrows and other tools customary for most Inuit populations in the region. The group had been struck by a pandemic in the 1820s that had taken the lives of the older and more knowledgeable members of the group. Losing the older members meant that the group also lost the necessary knowledge to create and use many of the tools and equipment the group were reliant on, since knowledge was passed along socially and not written down. Since the 1820s the population size had decreased. Many members of the group could remember these tools but they lacked the knowledge of how to create and use them. In 1862 another group of Inuit visited and reintroduced many of the tools they had lacked since the pandemic and the tools along with the necessary knowledge of usage and creation was quickly adopted by the group. The reintroduction of the tools and the necessary know-how allowed the population size to increase again (Mary-Rousselière, 1991; Rasmusen, 1908; as cited in Boyd et al., 2011).

What Tomasello et al. (1993) calls the “ratchet effect” refers to events where a specific tool has been preserved within a population to later be modified or improved, “ratcheting it up”.

One generation might create a helpful tool that allows them to better exploit their local environment and the following generation might keep the tool as it is, while later generations make changes and modifications. The "new" and modified version of the tool will then persist in a population and be taught to the coming generations until a new modification or improvement is made, and so on. The ratchet effect relies on high-fidelity cultural transmissions, which allows knowledge to be preserved, further taught and later modified and refined (Tennie et al., 2009; Tomasello et al., 1993). The ratchet effect explains modifications as shifts from one stable condition to another (Tomasello et al., 1993). A supplement, or challenger, to the ratchet effect is the "mountaineering analogy" proposed by Lombard (2016). Disparities in local technological change and variability within and between cultures represents a trend towards regional differentiations (Kuhn, 2006). The mountaineering analogy aims to explain a more dynamic approach where the development of technology is not linear but multidimensional and multidirectional. The difficulties of climbing a mountain, with ups and downs, usage of different anchoring points and aiming towards local high-points rather than constantly aiming for the global high represents the variability of tool modifications across cultures (Lombard, 2016).

Tomasello et al. (2005) proposes shared intentionality to be the crucial difference between human cognition and that of other animals. Shared intentionality means to participate in collaborative activities with other people, acting towards a mutual goal or goal-state. This is a strong form of intention-reading of others, and a form of cultural learning which motivates a sharing of mental states with others and using cognitive representations to do so. Mindreading, like theory of mind and shared intentionality, makes cultural transmissions easier. It is especially beneficial in contexts of teaching, but also important for interpersonal relationships and understanding of others. Culture reinforces mind reading through publicly shared psychological frameworks and is further linked with language by expressing beliefs, thoughts and desires through it (Whiten & Erdal, 2012). Cooperation and reciprocity support one another. Humans are more likely to cooperate if they expect to receive a reasonable amount of goodwill or material gain from the cooperation. Cooperation in turn supports cultural transmission, free sharing of information, knowledge and innovations which can lead to cultural evolution and culturally evolving technology (Whiten & Erdal, 2012).

Grammatical language is an adaptable tool for generation and exchange of thoughts and allows humans to store, elaborate and communicate ideas, insights, interpretations and knowledge (Bender & Beller, 2019). Language enhances several other abilities by helping to develop causal beliefs and incorporating them systematically to construct “theory-like representations”, which have been suggested to guide certain causal reasoning in humans (Penn & Povinelli, 2007). Through the integration of new knowledge and beliefs into larger explanatory frameworks language provides the accumulation of causal knowledge (Bender & Beller, 2019). Sharing of beliefs is not only beneficial in teaching-learning settings but can also arise in other social settings and interactions. Language is supported by culture through the process of language acquisition and is the most reliable way to relay high-fidelity cultural transmissions. Transmissions are done through a range of diverse approaches. Instructional learning, story-telling and social interactions between members of a group are all examples of transmissions of culture and knowledge (Whiten & Erdal, 2012). The greater cultural and innovation complexity the greater need for high-fidelity cultural transmissions, more easily achieved through language than any other strategy (Lewis & Laland, 2012).

2.2 Causal cognition

Creating and gaining understanding of causal relationships is of major importance to our species. Starting from infancy, humans have a drive to explore and test their causal assumptions which contributes to developing understanding of the world (Bonawitz & Muentener, 2017). Causality is the relation between two events, where one is the cause and the other is the consequence or effect (Bender & Beller, 2019; but see also Hume, 1739). Cognition concerned with causality, that is causal cognition, refers to how causal relations are observed, learned and reasoned about. Furthermore, causal cognition concerns causal relations and how they are represented both mentally and in language, as well as how the representations are applied to solve problems and reach goals (Bender & Beller, 2019). Causal cognition allows us to make predictions of events based on observations and experience, to affect and control events in the world and to draw inferences about causes to effects, and vice versa. The ability to acquire causal understanding and apply it to solve problems, create innovations, make predictions and reason about it is highly advantageous (Gärdenfors & Lombard, 2017). Human causal cognition is so advantageous that it has been considered one of the main driving forces in human evolution, even claimed to be the “one cognitive competence that underlies all later human achievements” (Stuart-Fox, 2014, p. 249). Bender and Beller (2019) propose that human culture and its impact on human causal cognition is what has led to the range and diversity of causal cognition in humans and what differentiates it between humans and non-human species. The human brand of causal cognition has been shaped by human culture, through cultural transmissions of knowledge and beliefs that provide the content of much human causal cognition.

Lombard and Gärdenfors (2021) proposes four critical requirements to define the usefulness and application of a technical tool, with an extrapolation of the list proposed by McCormack et al. (2011): (a) a tool or object's physical traits, (b) the physical and mental traits of the tool user and those of his/her target or audience, (c) the causal (mechanical or perceptual) principles that connect these traits, and (d) how the tool user understands the underlying principles and relationships between these different aspects (Lombard & Gärdenfors, 2021, para. 4). While causal reasoning skills is an advantage allowing our species to predict effects in the world, innovate and explore - it does not make us omniscient to the workings of the world. Furthermore, being able to predict an effect to a cause does not always translate to correct conclusions, reasoning or knowledge. Observations of the world and physical events can lead to useful knowledge about causal relationships, but relying only on observation can lead to faulty or incomplete knowledge and understanding.

2.3 Social learning and teaching

Learning can be categorised into different forms, such as imitative learning, instructed learning, collaborative learning and trial-and-error learning (Tomasello, 2016). Imitative learning can be observed in infants from the age of around 6-8 months, and is a form of social learning where an individual learns through a social environment rather than relying solely on innate or internal resources. Instructed or collaborative learning requires a more sophisticated understanding of other agents than imitative learning, and involves coordinated and integrated perspective taking, respectively (Tomasello, 2016).

Social learning can be defined as learning influenced by observation of or interaction with another individual (Box, 1984; Poulin-Dubois & Brosseau-Liard, 2016). Social learning can be widely observed in other animals, but humans remain unprecedented in our reliance on information communicated by others - cultural transmissions. Many of our cognitive abilities have been proposed to rely on our social learning abilities (Tomasello, 2016; Tomasello et al., 1993). The ability to learn from each other lowers the cost of gaining necessary knowledge for local adaptations, necessary for survival, and is a beneficial advantage. Instead of having to figure everything out on their own through trial-and-error, one can observe others and learn what to do or not to do from others, which Boyd et al. (2011, para. 1) proposed to be “crucial for human ecological success”. Socially learned behaviour, such as tool use, language and cultural norms, is difficult - if not impossible - to learn on one's own and constitutes a critical feature of child development (Kuzyk et al., 2019; Wood et al., 2016). To be able to understand perspectives of others and to diagnose their lack of or abundance of knowledge combined with motivations to help others understand has been proposed to be an instrumental human characteristic for the perseverance and evolution of human culture (Tomasello, 2000).

Being able to discern which person is the most accomplished in a field and choosing to learn from and/or imitate the most successful teacher can lower the cost of learning (Boyd et al., 2011). Choosing to learn from the most competent teacher is called selective social learning. Research contributing to understanding of children’s social learning is essential for advances in our understanding of human development. Extensive research the past two decades have shown that children learn selectively from some individuals over others (Poulin-Dubois & Brosseau-Liard, 2016; Sabbagh & Koenig, 2013; Stenberg, 2009; Stenberg & Hagekull, 2007; Thompson et al., 2022). Studies on infants have indicated that from the age of 12 months children often choose to imitate or look to whom they perceive to be the best teacher around for directives and not necessarily their parent (Stenberg, 2009; Stenberg & Hagekull, 2007). To learn how to properly filter between bad and good knowledge, children actively engage in social learning strategies. These strategies enable differentiation between reliable and unreliable sources to learn from and to later selectively trust certain sources instead of others (Koenig & Sabbagh, 2013; Kuzyk et al., 2019). In studies by Kuzyk et al. (2019) and Brosseau-Liard and Poulin-Dubois (2014) infants have shown attunement to teachers' (sources of information) accuracy, age and confidence to help them determine from whom to learn. From an evolutionary perspective, it would be advantageous for young children to prefer teachers who signal success or competence, indicating that the behaviour of these individuals would be worth copying and learning (Boyd et al., 2011; Chudek et al., 2013).

Some skills are far too advanced to be learned only through social learning and require an interactive teacher (Tehrani & Riede, 2008). Högberg et al. (2015, p. 850) proposes three grades of teaching: (1) Helping and correcting: the teacher interferes with what the learner does; (2) Showing: the teacher draws attention and demonstrates something to the learner; (3) Explaining: the teacher clarifies by communicating concepts and making the learner perceive patterns. Teaching can be defined from an evolutionary perspective as "a form of cooperative behaviour which functions to promote learning in others" (Thornton & Raihani, 2008, p. 1823). For teaching to be favoured by selection, the results would have to be favourable for both parties of the activity. In teaching, the results are dependent on the student's response and learning ability. If the students do not learn there is no benefit to the teacher. Feedback is an important factor in learning and teaching. In imitative learning a learner, novice to a task, observes the teacher, an expert at the task at hand. In an active teaching environment the teacher is not only performing a task, but also judges and modifies (Premack, 2004). If a student is confused regarding the material or task, the teacher would receive confusion as feedback. To gain understanding of a student, a teacher has to modify their teaching approach. Recursive feedback is when a teacher can observe their student practice or use what they have been taught. Evidence from experimental studies using recursive feedback have indicated higher learning results when the teaching was done face to face, and lower results when the teaching was done through computers (Okita & Schwartz, 2013; Okita et al., 2013). Teaching also generated more lasting knowledge in the students than working on their own without a teacher (Okita et al., 2013). In a similar manner, studies have shown that working with peers, equally novice to the material or task, produces better results than working individually due to constant feedback and opportunity to reason collaboratively (Derex & Boyd, 2015).

3. Previous Research

3.1 Original experiment by Derex et al.

Derex et al. (2019) aimed to examine whether human tool-use and CTC is reliant on causal reasoning abilities and constructed an experiment where participants' task was to optimise a physical wheel system. All participants were western university students, from Western, Educated, Industrialised, Rich, and Democratic (WEIRD) societies (Heinrich et al., 2010), shown to have a poor initial understanding of wheel dynamics (Proffitt et al., 1990). The task was done by moving weights on each of the four wheel spokes and thereby maximising the speed gained by the wheel during its 1 m inclination along a track (as exemplified in Figure

1). For this, each participant had five trials. Two dimensions were relevant in the moving of the weights along each spoke and influential to the gained speed of the wheel; the moment of inertia and the centre of mass. Participants had no physical interaction with the wheel of any sort (the study was done through a computer and physical artefacts were controlled by the experimenters), nor any contact or social interaction with other participants. Participants were organised into chains of five where each chain position and participant represented a separate generation in a generational transfer chain. The experiment contained two separate treatments with differences in the transfer of knowledge. One group received the last two configurations, with the associated speed, from the former participant while the other group received a similar transmission with an added depiction of theory from the former participant. Each treatment included 14 chains of five, rendering a participant pool of 140.

After five trials to optimise the wheel, each participant completed a knowledge test. The test functioned as a way of examining participants' knowledge of the causal technology of the wheel. Participants were presented with ten pairs of preset wheels, each pair differing in either their moment of inertia or centre of mass. The task was to decide which of the two wheels in each pair would gain the fastest speed in a similar task to the one they had just performed. Options to answer were "1" (for wheel one), "2" (for wheel two) and "No difference" (which was always incorrect).

Derex et al. (2019) found that the wheel became progressively better across the five artificial "generations", but that no explicit increase in causal understanding followed the improvement of the technical artefact. No significant differences were found between the two treatments, with regards to wheel speed or knowledge. From the results of the theory-treatment, Derex et al. concluded that the theory-transfer constrained exploration and had negative effects on participants' causal understanding. The results also indicated that most participants in the theory-treatment produced theories related to only one of two dimensions (moment of inertia or centre of mass) and rarely included correct information regarding both dimensions. Derex et al. described a "compensation phenomenon" - a better understanding of one of the two dimensions resulted in a worse understanding of the other. Participants who received high scores on test questions regarding one of the dimensions most often received low scores on questions regarding the other. Results from the study indicated a corollary between producing wheels optimised in one of the two dimensions and high scores of understanding on questions of the same dimensional nature. Derex et al. concluded their results to be supported by the cultural niche hypothesis.

3.2 Replicational experiment by Osiurak et al.

Osiurak et al. (2021) conducted a partial replication of the experiment by Derex et al. (2019) to examine the validity of their results and their methodological approach. The results from Osiurak et al.'s (2021) study supported an opposite conclusion than that of Derex et al. (2019) - improvements of the physical wheel system correlated with an increased technological causal understanding. This conclusion indicated support for the technical reasoning hypothesis, that technical reasoning skills are important in gaining an understanding of and improving technical content. Osiurak et al. (2021) raised several concerns regarding the methodological approach and its limitations in drawing the conclusions made by Derex et al. (2019).

Osiurak et al. (2021) conducted two separate treatments where both experimental designs were similar to the one of Derex et al. (2019), with 14 chains of five participants and with a replica of the wheel system. The partial replication by Osiurak et al. (2021) differed from the experiment by Derex et al. (2019) in how the knowledge test was conducted. In their response Osiurak et al. (2021) reasoned that the alternative of "No difference" was a confounding variable in the conclusions from the test scores. The conclusion from Derex et al. (2019) was that production and inheritance of wheel configurations biassed towards one of two dimensions (centre of mass or inertia) accumulated a poorer understanding of wheels with characteristics of the lesser produced dimension. Osiurak et al. (2021) instead reasoned that the poorer - or perhaps un-improved - understanding was due to a methodological error; the possibility of choosing "No difference" in the knowledge test. A better understanding of one of the dimensions induces a bias towards the option "No difference" for wheels of the less well understood dimension, unbeknownst to all participants that this option was always incorrect. This reasoning led to the removal of the "No difference" option in the study by Osiurak et al. (2021). The knowledge test (labelled *Analogous test*) in the study by Osiurak et al. included 24 items, instead of ten, with four answer options each. In addition, an experimental treatment examining how participants would transfer the gained knowledge to novice, but similar, situations was conducted. In the experimental treatment participants performed an additional test (labelled *Transfer test*) similar to the analogous test, but with two weights on each spoke (eight weights in total) instead of one weight per spoke (four in total). This was done due to technical reasoning being analogical - acquired understanding can be transferred to different, yet similar, situations. Osiurak et al. hypothesised that an increased understanding would be seen in both the analogous test and the transfer test, which was correct based on the results from their study.

3.3 Social learning, teaching and causal reasoning abilities

Further examining the social aspects of tool usage and optimisation, Derex and Boyd (2015) conducted an experiment where participants were to create totem poles of different complexity through a combinatorial computer program. Derex and Boyd examined the effects of group-level sociality and human reasoning ability through six separate treatments in their experiment: (1) fully isolated individuals; (2) isolated individuals with partial social information; (3) groups of three with full social information; (4) groups of six with full social information; (5) isolated automated learning algorithms, generating random variation; (6) groups of six automated learning bots. Creation of the totem poles could be done through several paths of different complexities, where higher levels of combinations in tool creation and tool use would increase the complexity of the resulting totem. Experimental group (2) created remarkably more complex totems than group (1), while group (3) and group (4) outperformed group (2). Social interaction between individuals gradually increased the complexity of the performed outcome, dependent on the level of social information shared in the group. The experiment by Derex and Boyd (2015) provided empirical evidence of collaboration and people working in groups are able to develop artefacts and technological advancements far too advanced for any isolated individual to invent during similar conditions and restraints. This is consistent with the cultural niche hypothesis suggesting that human prosperity through technological cultural advancements is a result from cultural information accumulation and not solely as a product of enhanced cognitive abilities (Boyd et al., 2011). Furthermore, the bots in the experimental groups (5) and (6) performed significantly worse in every instance. No isolated bot managed to create a totem pole, and only 0.004% of the bots in group (6) managed to achieve a finished product - outperformed even by isolated humans by a factor of two. This result highlights the importance of the ability of guided reasoning, in comparison to random implementation of tools (Suddendorf & Corbalis, 2007).

In an experimental study by Putt et al. (2014) two groups of novices were tasked to create a stone tool resembling an early Archeulean hand-axe (see Appendix B3 for reference). Whereas one of the groups received instructions by a teacher in the form of demonstration along with verbal instructions (referred to as “verbal group”), the other group could only observe - learning by demonstration (referred to as demonstration group). The results of the study indicated an inferior performance by the group which received verbal instructions along with the demonstration. However, Gärdenfors and Högberg (2017) proposes a different conclusion, based on the results of the study by Putt et al. (2014). The verbal group had received instructions regarding the importance of platform preparation (a vital step of preparation in creating the hand-axe), whereas the demonstration group had not. Where the verbal group had produced stone tools less resemblant to the hand-axe target, they had spent far more time on setting up and preparing proper platforms. They had copied the teacher's actions, with the necessary steps in preparation, instead of copying the desired results. This indicates an understanding of the subgoals, hierarchies and purpose of the task at hand, an understanding not observed in the demonstration group. Although the verbal group produced less resemblant hand-axes they had a better understanding of the full grasp of the task. The lesser results of this group could be derived from the time constraints, seeing as proper preparation takes time and could be difficult in complex unfamiliar tasks. The demonstration group tried to copy the desired resulting product, instead of copying the correct sequence of actions, showing evidence of emulation. While the verbal group could grasp, and partially implement, the necessary complexity of the task and its importance to the technology, the demonstration group could not. Gärdenfors and Högberg (2017) argued that the results of the verbal group would significantly improve, and overcome the results by the demonstration group, should the duration of the experiment be prolonged.

Caldwell and Millen (2009) examined the effects of social interaction (imitation - observing the process, emulation - observing the results, and teaching) on cumulative learning. This was done through a micro-society experiment where participants, in chains of ten, were to create paper aeroplanes flying as far a distance as possible. Cumulative effects were studied in separate treatments, dependent on methods of observation, teaching or neither. Results from the experiment showed evidence of improvements in every instance, despite absence of teaching or imitation. Caldwell and Millen further discussed the importance of their findings in how human and nonhuman primates differ in their observational skills, where nonhuman primates have been thought to depend more on emulation than imitation or teaching (Tomasello et al., 1993). Their experiment showed that humans could develop a cumulative culture with emulation as the only possible learning method. The arguably low complexity of the task is discussed as a vital aspect of the conclusions made from the study, where Caldwell and Millen dissuade from a generalisation of their findings towards more complex tasks and general teaching/learning behaviours. They argue that more complex tasks and knowledge not dependent on technical artefacts would be less likely to evolve without advanced social interactions. They refer to previous research (Caldwell & Millen, 2008) showing great variation of social transmission within micro-societies, arguing that traditions requiring more complex forms design evidently require more complex measures of learning.

Further examining both causal reasoning abilities and the effects of social interactions Zwirner and Thornton (2015) developed an experiment where participants in micro-societies were to produce baskets able to carry as much rice as possible. Separate treatments of the experiment studied how asocial learning, imitation, emulation and teaching would affect the resulting cultural cumulation. The results of the study indicated no significant differences between the treatments regarding carrying capacity of the manufactured baskets, while indications towards the teaching treatment producing superior baskets was visible but not statistically significant. Apart from this slight indication, participants in the teaching treatment, in which subjects received teaching from the previous generations, showed a great increase in the durability of their produced baskets. Durability in tools and artefacts would reasonably provide substantial advantages in nature, enabling long-term usage and a consistent source from which to learn and improve. Since no significant main effects were found between the treatments, in respect to carrying capacity, Zwirner and Thornton consider these results to be evidential support against the common belief of teaching being necessary for the cumulation of a complex technical culture (e.g. Tomasello et al., 1993). Instead, they reason that a complex causal understanding, to predict and represent causal outcomes, and the ability to reverse-engineer is in some instances enough, while highlighting the necessity of individual practice and implementation of knowledge.

The role of teaching, social interactions and causal understanding abilities in the evolution of human tool use are highly debated, as explained in the above paragraphs. While findings from studies by Zwirner and Thornton (2015) and Derex et al. (2019) puts focus mainly on the causal reasoning abilities of *Homo sapiens* and the ability to reverse engineer, studies by

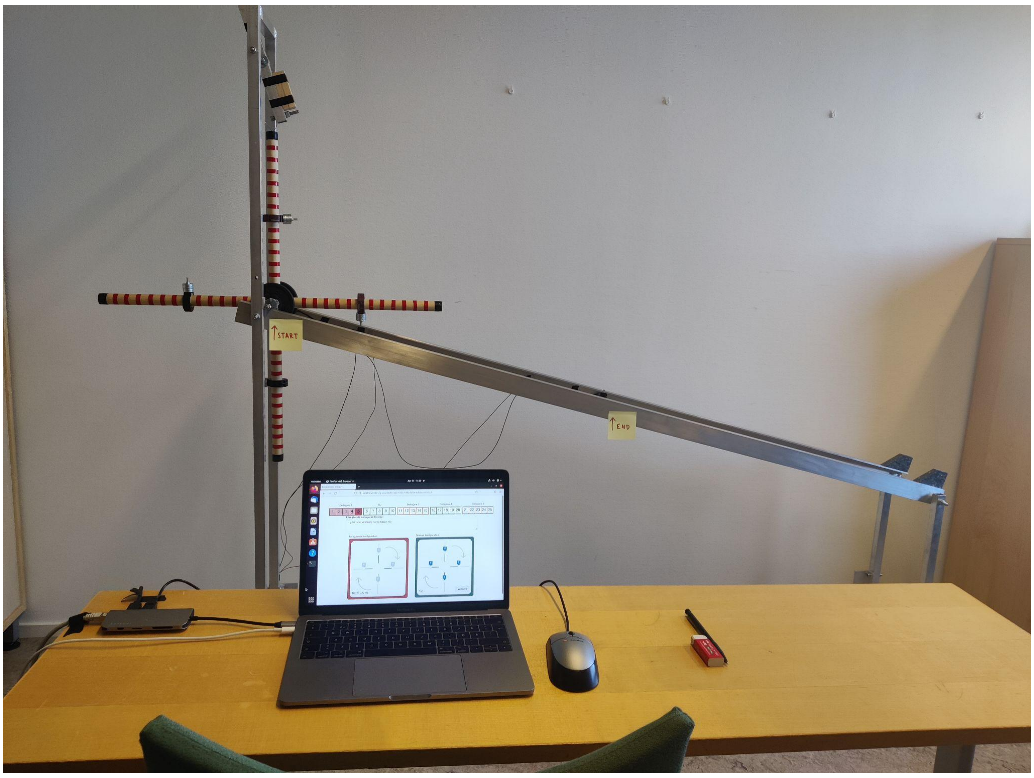
Gärdenfors and Högberg (2017) and Derex and Boyd (2015) highlights the importance of social interactions through teaching and learning from peers or generations before. The discussion by Gärdenfors and Högberg (2017), based on the study by Putt et al. (2014), gives rise to an interesting debate regarding the complexity of the tool and/or task. Imitation or emulation can be effective as a way of reaching a similar result, while verbal teaching is necessary in situations requiring a more complex understanding of the tool or procedure of manufacturing. Results from the aforementioned studies lie well with the mountaineering analogy proposed by Lombard (2016), indicating local highs of accumulation and differing complexities depending on the local norm of cumulation. If an individual is introduced to a more complex task with an accompanying theory, they expect to produce a more complex product or tool. If an individual instead has a more simple task introduced with no theory or prior knowledge, there are less expectations and more room to explore through trial and error, though without instructions or prior knowledge it would take far more practising to produce the more complex tool (Gärdenfors & Högberg, 2017).

4. Method

This study was conducted as means of examining the results from the previous experiment by Derex et al. (2019), both partially replicating the design of the original experiment and introducing new conditions to the existing experimental design. Participants first completed an experimental phase of altering the weight disposition of a physical wheel travelling down an inclined track (see Figure 1). The wheel dynamics depended on two theoretical dimensions - inertia and centre of mass - which could be manipulated by the participants' placements of weights. This stage was constructed through three separate conditions (see 4.5 Conditions). Participants then performed a knowledge test to assess the foundation of knowledge created through the course of the experiment (see 4.4.2 Knowledge test). The results were analysed and further discussed with regards to the conclusions and implications stated by Derex et al. (2019) and Osiurak et al. (2021) (see 6 Discussion and Appendix A1).

Figure 1

The experimental setup.



Note. The wheel was to roll down an inclined track from the position displayed in the figure. The notes on the track marked the distance relevant for time tracking. Participants chose their desired weights configurations through the displayed interface on the computer.

4.1 Participants

90 participants (47 male, 43 female), varying in age from 19 to 47 (M = 25.96, one participant did not provide age, SD = 5.7), took part in the study. All participants were connected to the University of Gothenburg (students (n = 84), faculty staff (n = 6)) and had full understanding of the Swedish language. Participants were split equally over three experimental groups rendering a population of n = 30 for each condition. Participants were recruited via social media platforms and through separate local communication channels within the University of Gothenburg. Informed consent was obtained from all participants in the study.

4.2 Experimental design

The study is of a Multilevel Mixed Model Design (Magezi, 2015), suitable for a hierarchical within-group structure. There are two variables, giving rise to a 3 x 5 model where condition and position in the chain act as independent variables. However, this would not fully represent the distinguished separations of the sought effects and the presentable variables. Neither is a subject's position in the chain regarded fully as an independent variable, it is merely an effect of the design of the experiment, although it represents differences in the experience from subjects' participation. Due to the complexity of the experimental design, three different effects were anticipated in this study. Firstly, a within-subject effect was hypothesised to emerge even though subjects only participated in one condition each and were only tested on one of the set independent variables. The within-subjects effect was sought in the development of performance between each participant's five trials and was prior to data collection thought of as a learning effect of each participant. This within-subjects effect was not of interest in the data analysis. However, participants were part of ordered chains within each condition. Every chain contained five subjects in a numbered order, from one to five. Each position in the chain figured in a slightly different manner and the results were analysed separately. This gave rise to the hypothesis of a second effect - a between-groups effect in the relation between the positions in the chain. Lastly, a between-groups effect was sought between the three separate conditions created, each with an alteration of the independent variables (see 4.5 Conditions).

4.3 Dependent variables of interest

Three dependent variables were measured; *Wheel speed*, *Test score* and *Theory score*. While the term "wheel time" is used in the experiment and rendered to participants, wheel speed is the actual dependent variable (see Appendix A1 for description). Wheel speed was recorded in the experimental phase of each condition and refers to the optimisation of the wheel. All five trials were recorded and saved individually for each participant. Wheel speed was measured in metres per hour (m/h).

Test scores refers to the number of correct answers displayed by participants on the knowledge test which followed the experimental phase. The scores were collected individually for each participant. Every correct answer rendered one point and the range of possible scores were 0 (zero, minimum) to 10 (ten, maximum).

Theory scores were quantified subsequently to the collection of data. Participants' suggestions of theory, transmitted to the adjacent participant, were scored independently with regards to the quality of information given based on the two affective dimensions of the wheel (moment of inertia and centre of mass). A score was given in the range of 0 - 2 for each dimension based on these criteria: *0 = no correct information; 1= partially correct information; 2 = full disclosure of correct information*. The separate scores of the two dimensions were then combined to a total score ranging from 0 - 4. Grading of the theory scores were performed blind to treatment and generation. All dependent variables were measured and collected separately.

4.4 Procedure

Participants were seated at a table placed 1 m from the experimental apparatus, facing the apparatus directly (see Figure 1). Each participant completed a consent form before starting the trials, containing information regarding the collection of demographic data and a non-disclosure agreement. Participants were assigned to a condition, a chain of five and a position in the chain depending on when they entered the study. Written instructions and summaries were given in Swedish to avoid any linguistic obstacles in the understanding of the task (see Appendix D1 & D2). Participants received instructions and summaries differing slightly depending on assigned condition and chain order. The order of the conditions were randomised and followed an ABCCBA-principle. Data collection was not performed blind to the conditions of the experiment.

After the full course of the experiment participants were asked questions related to demography (age, gender, level of education) and native tongue. Participants were also asked whether they had any higher level of academic background in physics or mechanical engineering. No segregation was made with regards to age, gender, native tongue or level of education. No personal data was applied to any stage of the analysis, nor was any participant excluded from the analysis due to their demography or background. Instead, the collection of demographic data was done to ensure any future studies where the given information might be of interest could be done with full sets of data. Full completion of the study was rewarded with a movie ticket.

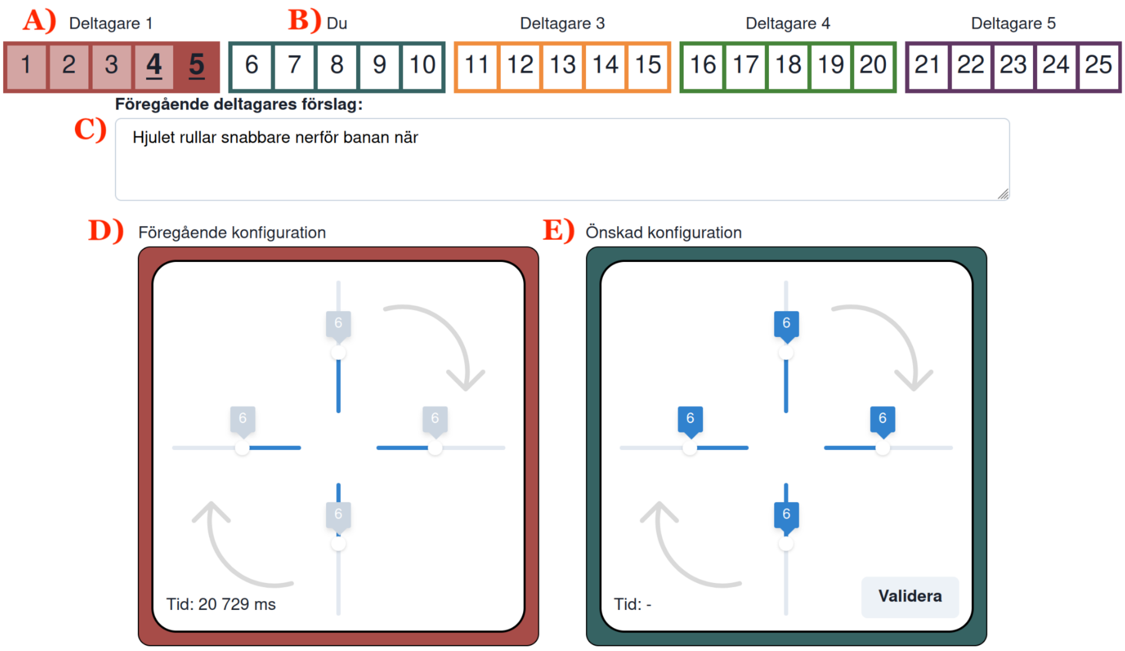
4.4.1 The experimental phase

In the experimental phase the task was for participants to minimise the time it took a wheel to cover 1 m on an inclined track (henceforth referred to as wheel time) (see Figure 1). Each participant conducted five trials to try and minimise the wheel time by moving a weight along each spoke of the wheel. Weights could be placed on one of 12 discrete positions along each of the four spokes, allowing 20736 possible unique configurations. Participants chose their desired configurations through a computer program where each spoke was portrayed along with their corresponding moveable weight with the possibility to slide each weight along the spokes (see Figure 2). Participants then had to validate their confirmation before the experimenter adjusted the weights on the physical wheel accordingly. The wheel was held in place by a mechanical lever as the weights were altered. The experimenter then lifted the lever and the wheel was released to roll down the tracks. The wheel time was recorded automatically by the computer program based on inputs from two time switches on the track, one at the release-point and one at the marked 1 m point of the track (see Appendix C1 & C2). While timekeeping only regarded the first metre (half of the rail's full length) - information fully disclosed to participants - the wheel continued to roll down to the bottom of the track, fully visible for the participants. Participants were given instant feedback on their wheel time and the associated configuration. During the full course of the experimental phase participants had the opportunity to view their last two configurations with corresponding wheel times.

Participants could also view configurations made by the former participant of the chain (see Figure 2), with slight variation depending on the participant's assigned experimental condition. Along with the former participant's configurations and wheel times a short suggestion of theory was included in the transfer between participants (see Figure 2 and 4.4.3 Theory transmissions). Viewing of a former participants' configurations was not possible for the first participant of each chain. There were no time limits for consultation or setting of configurations. Participants were informed that two of their configurations, depending on the assigned experimental group, would be transferred to the next participant in all cases but for the last participant of each chain since they would have no successor.

Figure 2

The Interface



Note. A) marks the previous participant, where configuration 4 and 5 is available. B) marks the current participant and their five trials. C) marks the previous participant's theory suggestions. D) marks the position where previous participants' configurations were viewable (now viewing participant ones last configuration). The configurations were displayed with corresponding times. E) marks where the current participant were to choose their desired configuration.

4.4.2 Knowledge test

Completion of the experimental phase introduced participants to the next part of the study - the knowledge test. This phase of the experiment was conducted to assess participants' understanding of the causality of the physical principles affecting the wheel. The knowledge test was conducted equally by all participants, with no regards to assigned condition or position in the chain. Participants were given a questionnaire along with instructions for the task (see Appendix D - Figure D1). A total of ten wheel pairs were displayed, one at a time, through a separate computer screen. The task was to assess each pair of wheels and decide which of the wheels would produce the fastest time in a task similar to the one they had just completed. Participants were free to switch between the wheel-pairs as they pleased and had no time constraints. The configurations of the wheel-pairs were directly copied from Derex et al. (2019), along with the correct answer key, to ensure replicational validity. Slight alterations were made to the visual design of the test as to fit the design of the physical wheel and the interface used in this study. In five pairs the wheels varied in their moment of inertia.

In the other five, wheels varied in the position of their centre of mass. Participants had to determine one of the wheels as faster than the other for each of the ten wheel pairs. Requests to answer “no difference” or “I don´t know” were declined by the experimenter, since this was an incorrect answer in every instance (see 3.2 Previous research & discussion in Osiurak et al., 2021). Participants were instead once again urged to choose between the two wheels in the pair. Data from participants who made such requests were included in the analysis.

4.4.3 Theory transmission

After completion of the knowledge test participants were to write a small suggestion of theory for the next participant in the chain. Instructions for the theory suggestion were given as a part of the same interface used in the experimental phase with equal opportunities to view past configurations. The suggestion always started with “The wheel covers the distance faster when…” (in Swedish: *Hjulet rullar snabbare nerför banan när…*) and had an upper limit of 340 characters in accordance with the original experiment (Derex et al., 2019). A lower limit to three words was set, with a blockage to submit if this was not fulfilled. The depiction of theory from the predecessor was constantly displayed in the interface throughout all the trials of every participant but for the first of each chain (see Figure 2).

4.5 Conditions

Three conditions were designed, each run in a separate treatment, all following the procedure described in *4.4.1 The experimental phase*, with slight variations between the different conditions (see Table 1). The conditions differed in the experimental phases, while the knowledge test and the theory transmission remained unchanged throughout. All conditions processed the same dependent variables.

Table 1

Compilation of Experimental Treatments

	Group A	**Group B**	**Group C**
Condition	Control	Verbalization	Best-two
Independent variables	The last two configurations were transferred to the following participant in each chain.	Participants were asked two questions regarding the effect of their manipulations to the weight distribution of the wheel. Included the same transfer of knowledge as the control group.	Differed in the transfer of knowledge. The *best* two, rather than the *last* two, configurations were transferred to the following participant.

4.5.1 Control group (group A)

Participants of the first condition, henceforth referred to as group A, acted as a control group. Group A performed an experiment replicated from the experimental group from the study performed by Derex et al. (2019) (see Table 1). Participants of group A inherited the last two configurations from the participant before them in the chain of five and were all informed about the nature of the transfer. The configurations were displayed in the interface as the participant entered the experimental condition, and were explained to be the former participants' last two trials with corresponding wheel time (see Figure 2). The transfer of configurations was accessible during the full course of the experiment. Participants had full disclosure to which of their trials would be transferred to the following participants.

4.5.2 Experimental group - Verbalization (group B)

Participants of group B performed an experiment identical to those of group A (see Table 1) apart from an added vocal, elaborating and interactive element labelled "Verbalization". After the second and fourth trials of each participant the experimenter asked a question referring to the participant's ability to reason. After the second trial, the experimenter asked "What was your process of reasoning when you placed the weights in this way?*"* (in Swedish: *Hur* *funderade du när du placerade vikterna på detta sätt?*). After the fourth trial the subject was asked “How do you think the placement of the weights affects the movement of the wheel?” (in Swedish: *Hur tror du att placeringen av vikterna påverkar hjulets rörelse?)*. All verbal communication occuring after the first question was asked up until completion of the experimental phase was recorded and later transcribed. The questions were not interfering with the procedure of the experiment and participants followed the same instructions as participants in group A.

4.5.3 Experimental group - Best-two (group C)

The second experimental group, henceforth referred to as group C, performed an experiment differing from group A and B in the transference of knowledge (see Table 1). Participants of group C were transferred the *best* two configurations and corresponding times from the former participant in the chain, instead of the *last* two as in groups A and B. Participants had full disclosure to which of their trials would be transferred to the following participants and were all informed about the nature of the transfer along with the viewable configurations. Participants followed the same instructions as participants in Group A and B apart from the information about transmission of the two best trials rather than the two last.

5. Results

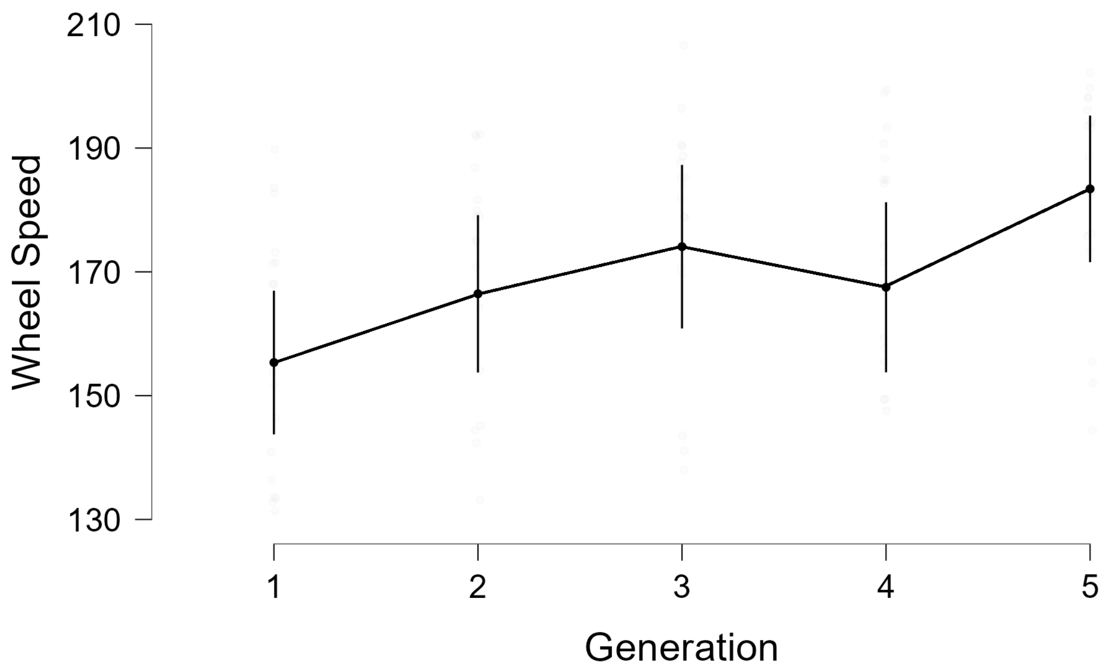
5.1 Wheel speed - All trials

Model 1: a Linear Mixed Model (LMM) was set with wheel speed (all five trials for each participant) as a dependent variable, generation as fixed effects variable and chain and participants as random effects grouping factors. A statistically significant effect was found on the generation variable ($F(4, 31.39) = [3.445]$, $p = .019$). A post hoc contrast test found a significant contrast between the first and the fifth generation (*Est.* = - 28.057, *SE* = 7.968, $z = -3.521$, $p < .001$, $p_{corrected} < .001$). The estimated marginal means for wheel speeds of the first generation (*EMM* =155.353, *SE* = 5.440, 95% *(CI)* = [144.691, 166.016]) was lower than the fifth (*EMM* =183.410, *SE* = 5.683, 95% *(CI)* = [172.271, 194.549]) (see Figure 3).

Model 2: a LMM set with wheel speed as a dependent variable, treatments as fixed effects variable and chain and participants as random effects grouping factors found no significant effects (p = .883) (see Appendix B - Figure B1 for visualisation).

Figure 3

Model 1 - The Average Wheel Speed across Generations



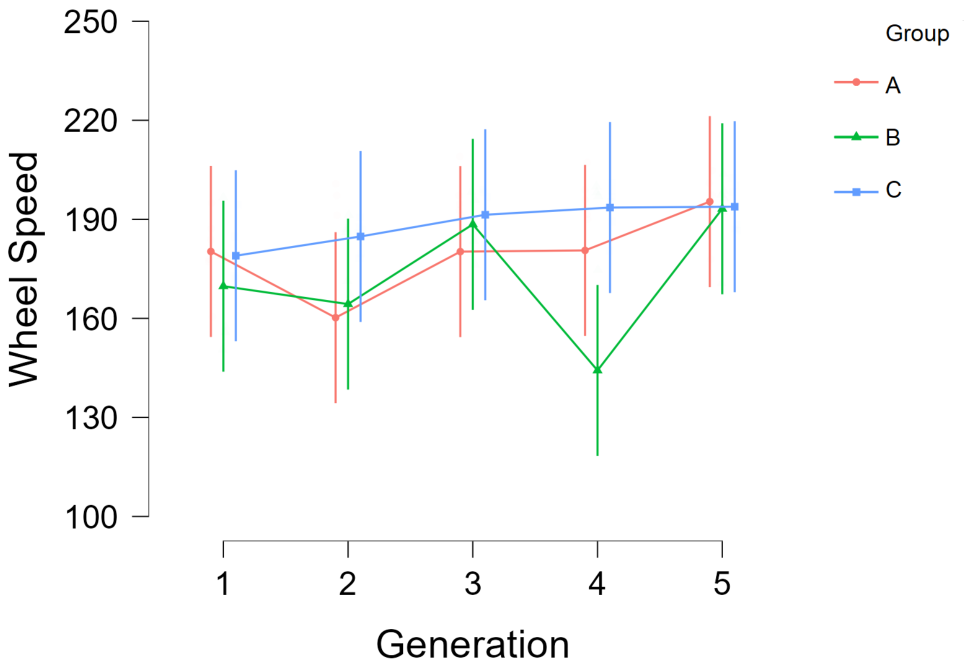
Note. The average wheel speeds across generations (including all five trials, by all participants). Wheel speed was measured in m/h. Error bars: 95% *(CI)*.

5.2 Wheel speed - Transferred trials

Model 3: a LMM was set with wheel speed of the transferred trials as a dependent variable, treatment as fixed effects variable and chain as random effects grouping factor. While no statistically significant effect was found on treatment (p = .220), the transfers of group C displayed a higher estimated marginal mean speed (*EMM* = 188.514, *SE* = 5.268, 95% *(CI)* = [175.997, 201.031]). Group A had the second highest speed (*EMM* = 179.331, *SE* = 5.268, 95% *(CI)* = [166.814, 191.848]) and group B had the lowest (*EMM* = 171.995, *SE* = 5.268, 95% *(CI)* = [159.997, 184.512]) (see Figure 4).

Figure 4

Model 3 - The Average Wheel Speed of the Transferred Trials between Treatments (groups).



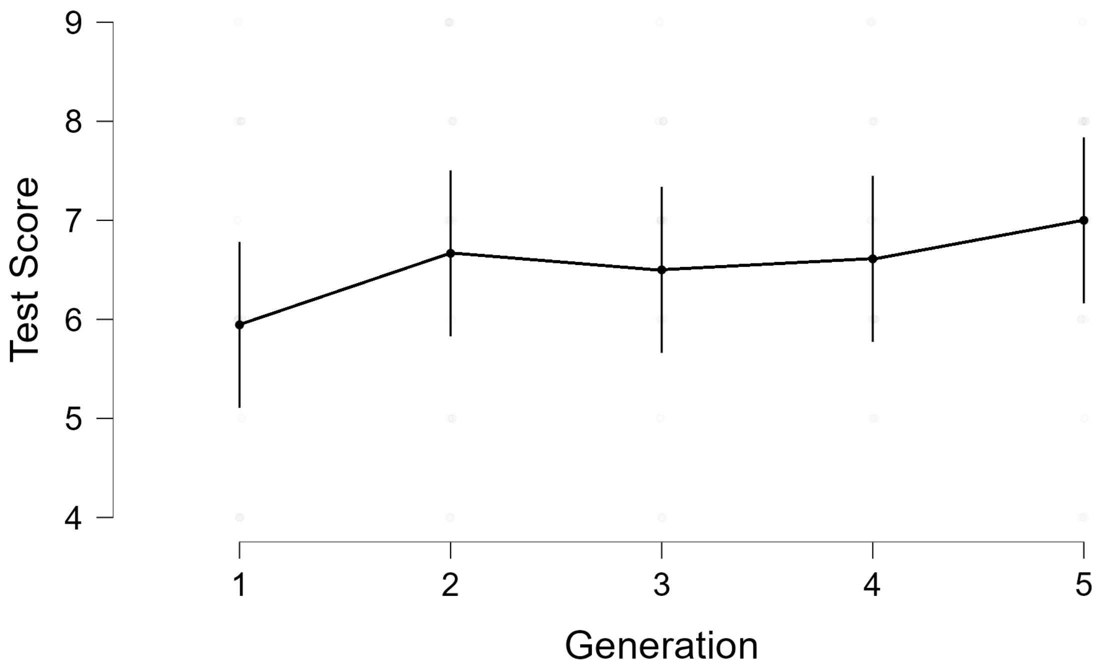
Note. The average wheel speed of the transferred trials across generations. The transferred trials included the last two trials for participants of group A and B and the best two trials for participants of group C. Wheel speed was measured in m/h. Error bars: 95% *(CI)*.

5.3 Test scores

Model 4: a LMM set with test scores as dependent variable, generation as fixed effects variable and chain as random effects grouping factor found no significant effect (p = .477). Although not statistically significant, there was a higher estimated marginal mean score for fifth generation participants' understanding (*EMM* = 7.000, *SE* = 0.421, 95% *(CI)* = [6.174, 7.826]) compared to first generation participants' (*EMM* =5.944, *SE* = 0.421, 95% *(CI)* = [5.118, 6.770]) (see Figure 5).

Figure 5

Model 4 - The Average of Test Scores on the Knowledge Test across Generations.

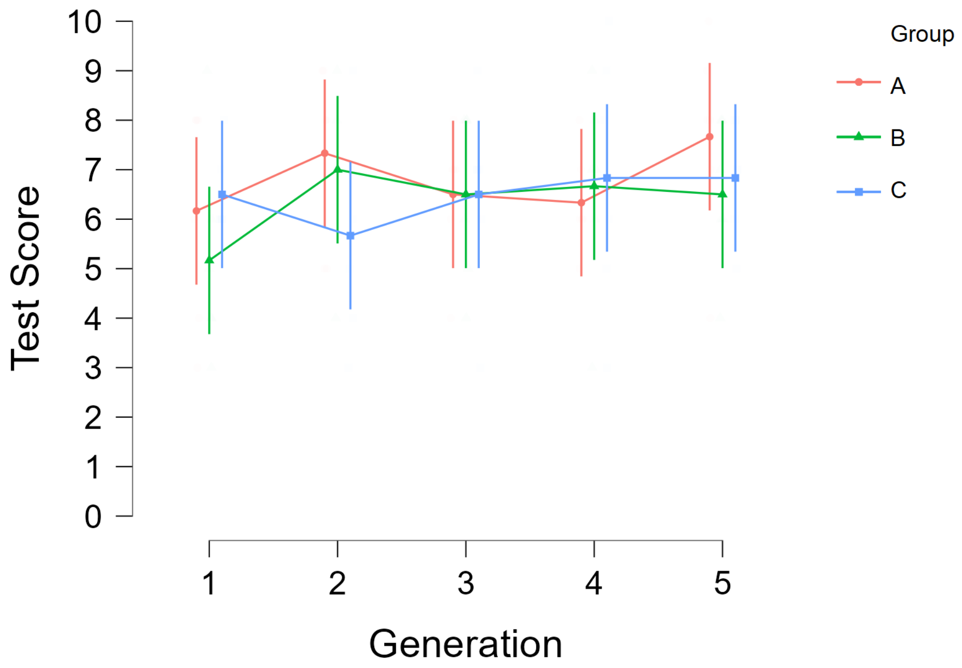


Note. The average test scores of all participants on the knowledge test across generations. The highest possible score on the knowledge test was 10 and the lowest was 0. Error bars: 95% *(CI)*.

Model 5: a LMM set with test scores as dependent variable, treatment as fixed effects variable and chain as random effects grouping factor found no significant effect (*p* = .707). Group A had a higher mean of test scores (*EMM* = 6.800, *SE* = 0.381, 95% *(CI)* = [6.054, 7.546]) than the other two groups, B (*EMM* = 6.367, *SE* = 0.381, 95% *(CI)* = [5.620, 7.113]) and C (*EMM* =6.467, *SE* = 0.381, 95% *(CI)* = [5.720, 7.213]) (see Figure 6).

Figure 6

Model 5 - The Average of Test Scores on the Knowledge Test between Treatments (groups).



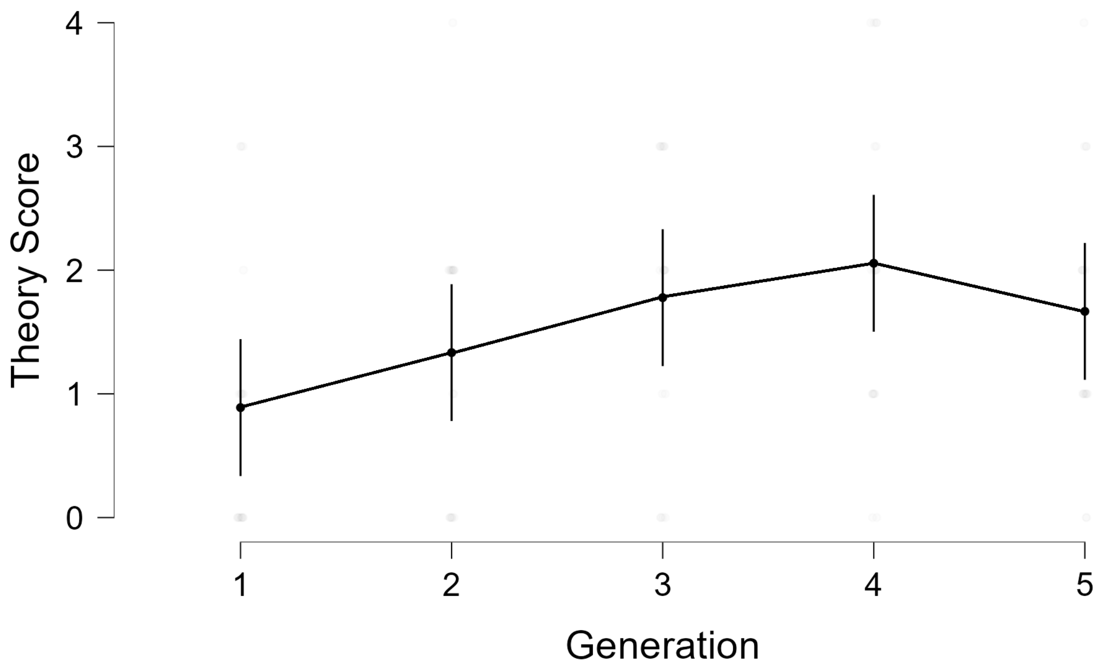
Note. The average test scores of all participants on the knowledge test across generations, separated by treatments. The highest possible score on the knowledge test was 10 and the lowest was 0. Error bars: 95% *(CI)*.

5.4 Theory transmissions

Model 6: a LMM set with the score of the graded theory transmissions as dependent variable, generation as fixed effects variable and chain as random effects grouping factor found a significant effect on the generations variable ($F(4, 68) = [3.279]$, $p = .016$). A post hoc contrast test found a significant contrast between the first and the third generation (*Est.* = -0.889, *SE* = 0.350, $z = -2.537$, $p = .011$, $p_{corrected} = .034$) and the first and the fourth generation (*Est.* = -1.167, *SE* = 0.350, $z = -3.330$, $p = .003$, $p_{corrected} < .001$). Comparison between the first and the fifth generation had a significant value for p without correction, but not when using Holm's method (*Est.* = -1.167, *SE* = 0.350, $z = -3.330$, $p = .026$, $p_{corrected} = .053$) and allowed no rejection of H_0. The estimated marginal means of score on the theory transmissions for the first generation was the lowest (*EMM* = 0.889, *SE* = 0.278, 95% *(CI)* = [0.345, 1.433]), followed by the third generation (*EMM* = 1.778, *SE* = 0.278, 95% *(CI)* = [1.234, 2.322]) while the fourth generation had the highest (*EMM* = 2.056, *SE* = 0.278, 95% *(CI)* = [1.512, 2.599]) (see Figure 7).

Figure 7

Model 6 - The Average of the Graded Theory Transmissions across Generations

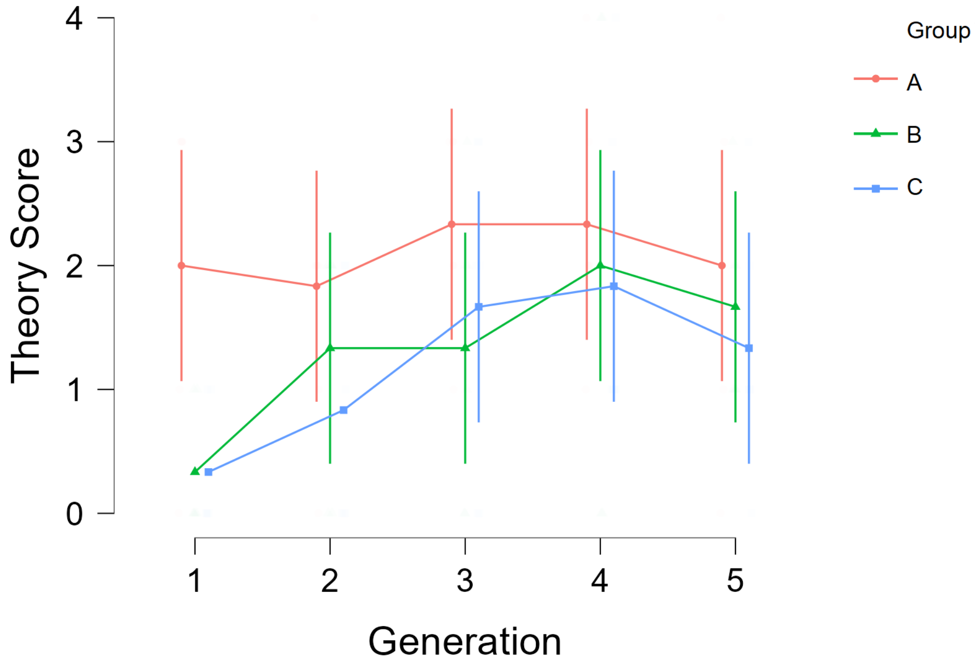


Note. The average scores of the graded theory transmissions of all participants across generations. The theory transmissions were graded on a score of 0-4. Error bars: 95% *(CI)*.

Model 7: a LMM set with the score of the graded theory transmissions as dependent variable, treatments as fixed effects variable and chain as random effects grouping factor found a significant effect of the treatment variable ($F(2, 15) = [3.716]$, $p = .049$). A post hoc contrast test comparing all treatments found a significant contrast between group A and C (*Est.* = -0.900, *SE* = 0.356, z = -2.526, $p = .012$, $p_{corrected} = .035$). The estimated marginal means of the graded theory transmission scores was the highest in group A (*EMM* = 2.100, *SE* =0.252, 95% *(CI)* = [1.606, 2.594]), followed by group B (*EMM* =1.333, *SE* = 0.252, 95% *(CI)* = [0.840, 1.827]) while group C had the lowest (*EMM* = 1.200, *SE* = 0.252 95% *(CI)* = [0.706, 1.694]) (see Figure 8).

Figure 8

Model 7 - The Average of the Graded Theory Transmissions between Treatments (groups).



Note. The average scores of the graded theory transmissions of all participants across generations, separated by treatments. The theory transmissions were graded on a score of 0-4. Error bars: 95% *(CI)*.

5.5 Verbalizations

The verbalised reasoning in group B (*n* = 30) was categorised qualitatively after identifying several trends outside of direct replies regarding the wheel mechanisms. Nine participants

(30%) expressed a restraint from exploration due to a desire of repeating and transferring their so far best configurations to the next participant. 12 participants (40%) expressed trust in the received theory suggestion. Four participants (13%) expressed a wish for a verification that the received theory suggestion and configuration was correct and to what degree. Three participants (10%) expressed a desire for information regarding the physical properties of the apparatus and the wheel (see Appendix A3 for transcriptions).

6. Discussion

This study investigated whether optimisation of a wheel system was followed by an increased causal understanding of the wheel system. The experiment was done as a partial replication of a former study by Derex et al. (2019), with added variables as an attempt to further mimic a natural social learning model within the boundaries set for the original experiment. Based on similar analyses as made by Derex et al., the result from this study was in line with the original study. However, further analysis indicated an increase in causal understanding of the wheel system, in contrast to the result from the study by Derex et al. (2019). The added variables in this study provided no significant evidence for neither optimisation nor understanding. Limitations identified in the experimental design based on earlier studies using micro-society paradigms and investigating CTC (Derex & Boyd, 2015; Gärdenfors & Högberg, 2017, Zwirner & Thornton, 2015), especially with regards to causal reasoning or social learning, raised questions about the suitability of this experiment for investigating CTC adequately.

6.1 Discussion of the results

The low statistical significance in various analyses could be attributed to the small sample size. This was expected and a continuation of the study will allow higher significance and more robust results. If the continuation of the study renders data continuous with the present findings, there is a high likelihood of finding significance in expected effects. Even though statistical significance was not reached for all sought effects, indications and trends in the data still merit discussion. The results showed a significant effect of wheel speed increasing over generations, indicating an improved optimisation between the first and fifth generation. Transferred trials in group C had a higher estimated marginal mean speed compared to those of group A and B. Fifth generation participants had higher scores on the knowledge test than the other generations, although this was not statistically significant. The graded scores of theory transmissions showed an increase from the first to the third generation and from the first to the fourth generation. This increase could be additional evidence of increased theoretical understanding of the wheel system, complementary to the score on the knowledge test. While the results of an increase from generation one to five was not statistically significant, the significant p-value, without correction, and the still low corrected p-value could allow the result to act as an indication. Further, the fifth generation's low score could be attributed to the lack of incentive - participants in the fifth generation were aware they were last in line and that their theory suggestion would not be transferred. Given incentive it is likely that participants in the fifth generation would produce a higher standard of theory transmissions.

The fifth generation had the highest wheel speeds and the highest scores on the knowledge test, indicating increased understanding along with optimisation of the wheel. This is additionally supported by the increase in score of theory transmissions from the first to the fourth generation. Pre-registered hypothesis 1 predicted an increase of displayed knowledge trailing an increase of optimised wheels, as indicated by the results.

Even though the effect was not statistically significant, participants of group C had the highest means of wheel speed of the transferred trials. In contrast, participants of group A had significantly higher scores on the graded theory transmissions than the other groups. Group A also had the highest wheel speeds across all trials and higher scores on the knowledge test although these were not significant effects. The pre-registered hypothesis 2 predicted the introduced variables in groups B and C would affect the results. While there were indications of differences between group A and groups B and C, they were not conclusive enough to confirm hypothesis 2.

6.2 Critique towards the design of the experiment:

6.2.1 Limitations within the design

The motivation to replicate in the first place was gathered from the partially assumptious conclusions made from the study by Derex et al. (2019). For instance, they concluded the following: "As predicted by the cultural niche hypothesis, our experiment shows that highly optimised technologies can emerge from the accumulation of many improvements made across generations linked by cultural transmission, without the need for an accurate causal understanding of the system" (Derex et al., 2019, p. 448) and states “Bows and arrows, houses and kayaks” (Derex et al., 2019, p. 446) to be examples of such optimised tools. This conclusion merits discussion. The complexity of the task is problematic and limiting in several ways. While optimisation to the technological artefact was prominent in both the presented results and those of Derex et al., the nature of the design limits the possibility of what Tomasello et al. (1993) refers to as a “ratchet effect”. The inevitable ceiling of the experiment makes it impossible to optimise the wheel in a boundary-breaking manner - the wheel can only reach a definite maximum speed. For example, a knife can be sharpened and shape-optimised to achieve a higher complexity and more rewarding usage, but the technical knowledge applied to create a knife could possibly also be implemented into creating a pair of scissors. Similarly, a kayak with rows can be optimised with regards to smoother and more efficient rowing, but the addition of sails lies outside the borders of rowing-optimisation. While the wheel in the experiment by Derex et al. (2019) can be optimised to roll faster, the boundaries of the experimental design severely limits the possibilities of outside-the-box goal-oriented optimisation, leaving the optimisation more comparable with the sharpening of a knife than the implementation of old knowledge to a new situation. The development and optimisation of bows and arrows, or houses and kayaks, and the optimisation of the wheel system from these experiments is arguably not translatable or comparable.

6.2.2 Social learning critique

Another conclusion made by Derex et al. (2019, p. 446) is: “Here we show that a physical artefact becomes progressively optimised across generations of social learners in the absence of explicit causal understanding”. Does the experiment provide enough social information in the transmission between participants to provide evidence regarding social learning? Studies where participants work in groups or transmit information verbally (Derex & Boyd, 2017) showed less artificial social situations to be favourable in creation and optimisation of technical artefacts. To define participants in these studies as full “social learners” (Derex et al., 2019, p. 446) is not in accordance with former research within the field of social learning (e.g. Box, 1984; Poulin-Dubois & Brosseau-Liard, 2016; Tomasello et al., 1993). Participants could emulate and reverse-engineer the results from previous participants, but had no opportunity to engage in a discussion or reason regarding the results.

To examine cumulative culture of technical artefacts under doubtful social conditions and in the absence of teaching seems unfulfilling, since complex tool technology and usage is highly dependent on social learning and teaching (Kuzyk et al., 2019; Wood et al., 2016). Further, Gärdenfors and Högberg (2017) suggests that the results from Putt et al. (2014) provide an example of the effects of social transmissions of verbal teaching. Analysing the study by Derex et al. (2019) with the levels of teaching suggested by Högberg et al. (2015) in mind highlighted weaknesses in the methodological design and conclusions made by Derex et al. (2019). The social transmissions between generations enable only emulation - the availability of a final result to copy or reverse engineer. In Derex et al.s' (2019) experiment the theory depictions between generations are, supposedly, a means to enable imitation or an explanation by the teacher (former generation). Teaching, however, is in nature a two-way interaction. In a cumulative culture a teacher must modify their approach according to the feedback given by the student to achieve optimised results. Recursive feedback highly enhances the quality of the transfer of knowledge in a teaching situation (Okita & Schwartz, 2013; Okita et al., 2013). Likewise, all three levels of teaching proposed by Högberg et al. (2015) includes an interaction between the teacher(s) and the student(s). In the experiment by Derex et al. (2019), and as a result also in this experiment, no interaction between participants was allowed. The methods of transmittance between generations do not qualify as interactive teaching by the variations proposed by Högberg et al. (2015), but merely as emulation.

The independent variable of verbalisation, group B, was added to this study as a means to approach a more naturalistic situation with verbal social learning, within the boundaries set by Derex et al. (2019) for their experiment. While the verbalisation was not equal to a social context in the form of reasoning with a peer or a teacher, the aim was to introduce a variable related to the phenomena of a social culture. The aim was to encourage the participant to verbalise their reasoning to more closely mimic a social learning situation. A similar study, performed with social interaction in a more naturalistic setting or with groups of participants working together, would arguably produce a more robust result than displayed based on former studies (Derex & Boyd, 2015).

In nature it is often more experienced individuals taking on the role as a teacher. However, in this experiment, the teachers are equal to the learners, apart from having recently performed the five trials. Teachers are expected to teach after only 20 minutes of unguided and unvalidated learning by themselves. This is highly unnatural and since participants were aware of the chain functionality this introduces a question of validity for the transfers. In nature, selective social learning strategies are prominent. Each subject of learning can filter selectively, based on perceived success e.g., and differentiate between available teachers - choosing from whom to learn and who to trust (Boyd et al., 2011). The variable introduced in group C was set as an attempt to further mimic a form of selective learning lacking in the original experiment. Participants of group A or B had no means of validating the information they received from the former participant, possibly creating distrust. Participants in group C received what was certified as the two best attempts from the former participant which was hypothesised to cause a form of validation of the transfer. However, no participant had the option to select from whom to learn or copy - a factor hypothesised to have a larger effect on the results than the introduced variable.

Following the critique presented with regards to social environments, learning and teaching, micro-society paradigms might be considered inappropriate in the exploration of cumulative culture. Since culture by nature is social (Laland et al., 2011), an experiment with conditions deemed asocial, or insufficiently social, can not provide substantial conclusions regarding cumulative culture.

6.2.3 Causal reasoning critique

Based on two of the four requirements, critical to define the usefulness and application of a technical tool, proposed by Lombard and Gärdenfors (2021) the design of the experimental phase is inadequate. These two requirements of knowledge are: "(a) a tool or object's physical traits; (b) the physical and mental traits of the tool user and those of his/her target audience" (Lombard & Gärdenfors, 2021, para. 4). Participants had no physical interaction with the wheel whatsoever, consequently not allowing the weights to be fully understood as causal affectors but merely as an arbitrary cause of outcome. This would be directly limiting according to requirement (a). Several participants expressed a desire to know the actual weight of the moveable weight units (see Appendix A3 for transcriptions). Further, requirement (b) states the essentiality of a social environment in development of technical understanding. While Derex et al. (2019) claims to study social learners, the grounds on which they claim their experiment to be of social nature are not sufficient to express necessary information according to criteria (b) proposed by Lombard and Gärdenfors (2021). Further, the artificiality of the experiment is a major confounding for conclusions in the nature of cumulative culture - arguably even more important of a critique since Derex et al (2019) states that the wheel system was used due to participants (western university students) having a poor understanding of wheel dynamics (Proffitt et al., 1990). The three steps of alternation between conceptual realms in the experiment can arguably cause conceptual challenges of understanding and transmission of knowledge; first, weight alterations in a computer interface; secondly, observing a physical wheel; thirdly answering a test of knowledge in a novice environment; and lastly, to provide a written suggestion of theory. These challenges are supported by experimental studies by Okita and Schwartz (2013) and Okita et al. (2013) showing evidence of the difficulties in an artificial learning situation. With the arguments presented above in mind, claiming the knowledge test to evaluate participants' understanding of a causality regarding the physical tool seems inadequate. Rather, it would seem, the knowledge test refers to the ability to answer a test regarding artificial and hypothetical wheels.

6.3 Further research

The continuation of this study will reveal more data from which more robust and precise conclusions will be drawn. Besides this, further development of the experimental design to include more natural social learning conditions is suggested. To include groups of peers and expert teachers within the transmission of knowledge and an allowance for selective social learning is deemed critical for a base of which to draw disclosing conclusions regarding cumulative culture. At the very least, a physical social contact between participants would be an interesting introduction to inch closer to natural circumstances. Further, to develop a test of knowledge within the same area of implementation, not based on the transmittance of knowledge between dimensions, is encouraged. Furthermore, for future research to aim for a more inclusive approach and include participants outside of WEIRD societies is highly encouraged.

7. Conclusions

The results of this study show that optimisation of technical artefacts, wheel configurations in this study, is followed by an increased causal understanding of the artefact. Participants' ability to transfer more complex and correct theories in the later generations is no doubt a clear indication of the increased knowledge and understanding. The knowledge test provided similar indications although its validity is to be questioned with regards to relevance. However, the identified problems of the experimental design, originally created by Derex et al. (2019) and partially replicated in this study, show that the experiments do not necessarily explore cumulative technological culture. The usage of micro-society paradigms to explore cumulative culture needs to include more relevant aspects of natural and social conditions to be able to draw more correct conclusions regarding human cultural evolution. By excluding important aspects the conclusions are not directly transferable and entails a need to proceed with caution when doing so.

8. References

Bender, A., & Beller, S. (2019). The Cultural Fabric of Human Causal Cognition. *Perspectives on Psychological Science*, 14(6), 922-940. 10.1177/1745691619863055

Bonawitz, E., & Muentener, P. (2017). *The Development of Causal Reasoning*. In The Oxford Handbook of Causal Reasoning (Oxford Library of Psychology, pp. The Oxford Handbook of Causal Reasoning, 2017-06-30). Oxford University Press. 10.1093/oxfordhb/9780199399550.013.40

Boyd, R., & Richerson, P. (1985). *Culture and the evolutionary process.* Chicago: University of Chicago Press.

Boyd, R., Richerson, P., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences - PNAS*, 108(Supplement 2), 10918-10925. 10.1073/pnas.1100290108

Box, H. O. (1984). Social Ecology. In *Primate Behaviour and Social Ecology* (pp. 6-51). Springer, Dordrecht.

Brosseau-Liard, P., & Poulin-Dubois, D. (2014). Sensitivity to Confidence Cues Increases during the Second Year of Life. *Infancy*, 19(5), 461-475. 10.1111/infa.12056

Caldwell, C., & Millen, A. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior*, 29(3), 165-171. 10.1016/j.evolhumbehav.2007.12.001

Caldwell, C., & Millen, A. (2009). Social Learning Mechanisms and Cumulative Cultural Evolution: Is Imitation Necessary? *Psychological Science*, 20(12), 1478-1483. 10.1111/j.1467-9280.2009.02469.x

Chudek, M., Brosseau-Liard, P., Birch, S., & Henrich, J. (2013). Culture-Gene Coevolutionary Theory and Children's Selective Social Learning. In *Navigating the Social World* (pp. Navigating the Social World, 2013-04-23). New York: Oxford University Press. 10.1093/acprof:oso/9780199890712.003.0033

Derex, M., Bonnefon, J., Boyd, R., & Mesoudi, A. (2019). Causal understanding is not necessary for the improvement of culturally evolving technology. *Nature Human Behaviour*, 3(5), 446-452. 10.1038/s41562-019-0567-9

Derex, M., & Boyd, R. (2015). The foundations of the human cultural niche. *Nature Communications*, 6(1), 8398. 10.1038/ncomms9398

Gärdenfors, P., & Högberg, A. (2017). The Archaeology of Teaching and the Evolution of Homo docens. *Current Anthropology*, 58(2), 188-208. 10.1086/691178

Gärdenfors, P., & Lombard, M. (2020). Technology led to more abstract causal reasoning. *Biology & Philosophy, 35(4), Biology & philosophy*, 2020-07-06, Vol.35 (4). 10.1007/s10539-020-09757-z

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33(2-3), 45-67. 10.1017/S0140525X0999152X

Hume, D. (1739). *A treatise of human nature*. Oxford.

Hunt, G., & Gray, R. (2003). Diversification and cumulative evolution in New Caledonian crow tool manufacture. *Proceedings of the Royal Society B: Biological Sciences*, 270(1517), 867-874. 10.1098/rspb.2002.2302

Hunt, G., & Gray, R. (2004). The crafting of hook tools by wild New Caledonian crows. *Proceedings of the Royal Society B: Biological Sciences*, 271(Suppl 3), S88-S90. 10.1098/rsbl.2003.0085

Högberg, A., Gärdenfors, P., & Larsson, L. (2015). Knowing, Learning and Teaching—How Homo Became Docens. *Cambridge Archaeological Journal*, 25(4), 847-858. DOI: 10.1017/S0959774315000402

JASP Team .(2022). JASP (Version 0.16.1.) [*Computer software*]. Web. https://jasp-stats.org/

Kirby, S., & Tamariz, M. (2022). Cumulative cultural evolution, population structure and the origin of combinatoriality in human language. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 377(1843), 20200319. 10.1098/rstb.2020.0319

Koenig, M., & Sabbagh, M. (2013). Selective Social Learning. *Developmental Psychology*, 49(3), 399-403. 10.1037/a0031619

Kuhn, S. (2006). Trajectories of Change in the Middle Paleolithic of Italy. *In Transitions Before the Transition (Interdisciplinary Contributions To Archaeology*, pp. 109-120). Boston, MA: Springer US. 10.1007/0-387-24661-4_6

Laland, K., Atton, N., & Webster, M. (2011). From fish to fashion: Experimental and theoretical insights into the evolution of culture. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 366(1567), 958-968. 10.1098/rstb.2010.0328

Legare, C. (2017). Cumulative cultural learning. *Proceedings of the National Academy of Sciences* - PNAS, 114(30), 7877-7883. 10.1073/pnas.1620743114

Lewis, H., & Laland, K. (2012). Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society B*, 367(1599), 2171-2180. 10.1098/rstb.2012.0119

Lombard, M. (2016). Mountaineering or Ratcheting? Stone Age Hunting Weapons as Proxy for the Evolution of Human Technological, Behavioral and Cognitive Flexibility. *In The Nature of Culture (Vertebrate Paleobiology and Paleoanthropology*, pp. 135-146). Dordrecht: Springer Netherlands. 10.1007/978-94-017-7426-0_12

Lombard, M., & Gärdenfors, P. (2017). Tracking the evolution of causal cognition in humans. *Journal of Anthropological Sciences*, 95, 219-234. 10.4436/jass.95006

Lombard, M., & Gärdenfors, P. (2021). Causal Cognition and Theory of Mind in Evolutionary Cognitive Archaeology. *Biological Theory,* 2021-02-11. 10.1007/s13752-020-00372-5

Magezi, D. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, 6, 2. 10.3389/fpsyg.2015.00002

Mary-Rousselière, G. (1991). *Qitdlarssuaq, The Story of a Polar Migration*. Wuerz Publishing.

McCormack, T., Hoerl, C., & Butterfill, S. (2011). Tool Use and Causal Cognition (Consciousness & Self-Consciousness Series). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199571154.001.0001

Moll, H., Nichols, R., & Pueschel, E. (2020). Shared intentionality shapes humans' technical know-how. *The Behavioral and Brain Sciences*, 43, E172. 10.1017/S0140525X20000163

O'Brien, M., & Bentley, R. (2021). Genes, culture, and the human niche: An overview. *Evolutionary Anthropology*, 30(1), 40-49. 10.1002/evan.21865

Okita, S., & Schwartz, D. (2013). Learning by Teaching Human Pupils and Teachable Agents: The Importance of Recursive Feedback. *The Journal of the Learning Sciences*, 22(3), 375-412. 10.1080/10508406.2013.807263

Okita, S., Turkay, S., Kim, M., & Murai, Y. (2013). Learning by teaching with virtual peers and the effects of technological design choices on learning. *Computers and Education*, 63, 176-196. 10.1016/j.compedu.2012.12.005

Osiurak, F., De Oliveira, E., Navarro, J., & Reynaud, E. (2020). The Castaway Island: Distinct Roles of Theory of Mind and Technical Reasoning in Cumulative Technological Culture. *Journal of Experimental Psychology. General*, 149(1), 58-66. 10.1037/xge0000614

Osiurak, F., Lasserre, S., Arbanti, J., Brogniart, J., Bluet, A., Navarro, J., & Reynaud, E. (2021). Technical reasoning is important for cumulative technological culture. *Nature Human Behaviour*, 5(12), 1643-1651. 10.1038/s41562-021-01159-9

Osiurak, F., & Reynaud, E. (2020). The elephant in the room: What matters cognitively in cumulative technological culture. *The Behavioral and Brain Sciences*, 43, E156. 10.1017/S0140525X19003236

Osiurak, F., & Reynaud, E. (2020). The elephant in the China shop: When technical reasoning meets cumulative technological culture. *The Behavioral and Brain Sciences*, 43, E183. 10.1017/S0140525X20000291

Penn, D., & Povinelli, D. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology*, 58(1), 97-118.

Poulin-Dubois, D., & Brosseau-Liard, P. (2016). The Developmental Origins of Selective Social Learning. *Current Directions in Psychological Science : A Journal of the American Psychological Society*, 25(1), 60-64. 10.1177/0963721415613962

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences - PNAS*, 107(Supplement 2), 8993-8999. 10.1073/pnas.0914630107

Premack, D. (2004). Is Language the Key to Human Intelligence? *Science (American Association for the Advancement of Science)*, 303(5656), 318-320. 10.1126/science.1093993

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, 1(4), 515-526. 10.1017/S0140525X00076512

Proffitt, D., Kaiser, M., & Whelan, S. (1990). Understanding wheel dynamics. *Cognitive Psychology*, 22(3), 342-373. 10.1016/0010-0285(90)90007-Q

Putt, S., Woods, A., & Franciscus, R. (2014). THE ROLE OF VERBAL INTERACTION DURING EXPERIMENTAL BIFACIAL STONE TOOL MANUFACTURE. *Lithic Technology*, 39(2), 96-112. 10.1179/0197726114Z.00000000036

Rasmusen, K. (1908). *People of the Polar North.* London: Kegan Paul, Trench, Trubner and Co. Print

Richerson, P., & Boyd, R. (2005). *Not by genes alone : How culture transformed human evolution.* Chicago: University of Chicago Press.

Shea, J. (2017). Occasional, obligatory, and habitual stone tool use in hominin evolution. *Evolutionary Anthropology*, 26(5), 200-217. 10.1002/evan.21547

Stenberg, G. (2009). Selectivity in Infant Social Referencing. *Infancy*, 14(4), 457-473. 10.1080/15250000902994115

Stenberg, G., & Hagekull, B. (2007). Infant Looking Behavior in Ambiguous Situations: Social Referencing or Attachment Behavior? *Infancy*, 11(2), 111. 10.1080/15250000709336635

Stuart-Fox, M. (2014). The origins of causal cognition in early hominins. *Biology & Philosophy*, 30(2), 247-266. 10.1007/s10539-014-9462-y

Suddendorf, T., & Corballis, M. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *The Behavioral and Brain Sciences*, 30(3), 299-313. 10.1017/S0140525X07001975

Taylor, A., Miller, R., & Gray, R. (2012). New Caledonian crows reason about hidden causal agents. *Proceedings of the National Academy of Sciences - PNAS*, 109(40), 16389-16391. 10.1073/pnas.1208724109

Tehrani, J., & Riede, F. (2008). Towards an archaeology of pedagogy: Learning, teaching and the generation of material culture traditions. *World Archaeology*, 40(3), 316-331. 10.1080/00438240802261267

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B*, 364(1528), 2405-2415. 10.1098/rstb.2009.0052

Thompson, B., Van Opheusden, B., Sumers, T., & Griffiths, T. (2022). Complex cognitive algorithms preserved by selective social learning in experimental populations. *Science (American Association for the Advancement of Science)*, 376(6588), 95-98. 10.1126/science.abn0915

Thornton, A., & Raihani, N. (2008). The evolution of teaching. *Animal Behaviour*, 75(6), 1823-1836. 10.1016/j.anbehav.2007.12.014

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard university press.

Tomasello, M. (2000). Culture and Cognitive Development. *Current Directions in Psychological Science : A Journal of the American Psychological Society*, 9(2), 37-40. 10.1111/1467-8721.00056

Tomasello, M. (2016). The ontogeny of cultural learning. *Current Opinion in Psychology*, 8, 1-4. 10.1016/j.copsyc.2015.09.008

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *The Behavioral and Brain Sciences*, 28(5), 675-691. 10.1017/S0140525X05000129

Tomasello, M., Kruger, A., & Ratner, H. (1993). Cultural learning. *The Behavioral and Brain Sciences,* 16(3), 495-511. 10.1017/S0140525X0003123X

Tooby, J., & DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. *The evolution of human behavior: Primate models*, 183-237.

Whiten, A., & Erdal, D. (2012). The human socio-cognitive niche and its evolutionary origins. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 367(1599), 2119-2129. 10.1098/rstb.2012.0114

Whiten, A., & Van Schaik, C. (2007). The evolution of animal 'cultures' and social intelligence. *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 362(1480), 603-620. 10.1098/rstb.2006.1998

Wood, L., Harrison, R., Lucas, A., McGuigan, N., Burdett, E., & Whiten, A. (2016). "Model age-based" and "copy when uncertain" biases in children's social learning of a novel task. *Journal of Experimental Child Psychology*, 150, 272-284. 10.1016/j.jecp.2016.06.005

Yamamoto, S., Humle, T., & Tanaka, M. (2013). Basis for Cumulative Cultural Evolution in Chimpanzees: Social Learning of a More Efficient Tool-Use Technique. PloS One, 8(1), E55768. 10.1371/journal.pone.0055768

Zwirner, E., & Thornton, A. (2015). Cognitive requirements of cumulative culture: Teaching is useful but not essential. *Scientific Reports*, 5(1), 16781. 10.1038/srep16781

Appendix A

Data analysis, data and transcriptions

A1 - Data analysis

In the analysis of the data the measured wheel times were converted from milliseconds (ms) to average speed in metres per hour (m/h) using the formula ($x\,ms = 1 \div 360000\ m/h$). This was done to avoid the difficulty of attributing an arbitrary value of duration to wheel configurations that did not roll at all. Measuring in velocity gave the opportunity to attribute the value of 0 m/h (zero) to non-functional configurations to achieve a negative result. This corresponds well with the data of non-functioning wheels, which had no motion and hence no velocity either. Optimisation of wheel speed rendered a continuously lower score of duration, hence attributing the value of 0 (zero) ms to non-functioning wheel configurations would drastically improve the overall performance of the concerned participant. The option of attributing a "non-value" (a non-existing value, removing the data point completely from the analysis) was dismissed since the non-functioning wheel configurations were a part of the performance nonetheless.

The statistical program JASP version 0.16.1 (JASP Team, 2022) was used to carry out statistical analysis on the data gathered in the study. Linear Mixed Models was used due to the hierarchical structures of the experimental design with participants' different chain positions. Each model was fitted with a dependent variable, a fixed effects variable and either one or two random effects grouping factor(s). For the models which found a significant effect, contrasts were specified for multiple comparison testing to find affected differences. Multiple comparison statistical testing was adjusted for family wise error rate (FWER) using the Holm method, to avoid Type I errors (false positives).

A2 - Data

All analyses used and results given in this study can be found online through this link: https://osf.io/9qavw/. From the uploaded JASP files it is possible to see which statistical test was run, which input variables were used and the subsequent results. The uploaded CSV files include all raw data that was analysed. All transcriptions can be found in the uploaded excel spreadsheet, and a text file with supplementary information to understand the tags in which they were categorised in.

A3 - Transcriptions

All transcriptions can be found through the spreadsheet "Group B - Verbalised Reasoning" at https://osf.io/9qavw/. Explanation for the categorizations and the spreadsheet can be found through the associated text file. Selected transcriptions, referenced in discussion and results, are depicted and translated to English below.

Restrained from exploration
“I think that I will make another configuration like the one I did before”, Group B, Chain 4, Participant 4.

“Is it important when I am on trial ‘20’, that this is my final trial in some way? I will have to think about that, because that trial will transfer on to participant 5 and that I want ‘20’ to be - that I am sure of the information I give there. So, trial ‘18’ can still be used as an experimental trial”, Group B, Chain 6, Participant 4.

Selective Social Learning
“I don’t know what it is the former participant has gotten as results, if this is like the best or if this is just average or whatever it is”, Group B, Chain 3, Participant 3.

“So I can’t find out how fast it has gone, as a record? Hmm, what shall we do with this last configuration then? This configuration was the best so far, can it get faster than this?”, Group B, Chain 4, Participant 2

Reasoning regarding the physical wheel system
“I wonder how great the variation is, like say I would roll the same configuration 5 times, or say 10 times”, Group B, Chain 3, Participant 2.

“I would like to know how heavy they are. Are they heavy weights? Or, they might not weigh anything at all. It might not matter, or yes it does… They must weigh a little at least”, Group B, Chain 4, Participant 5

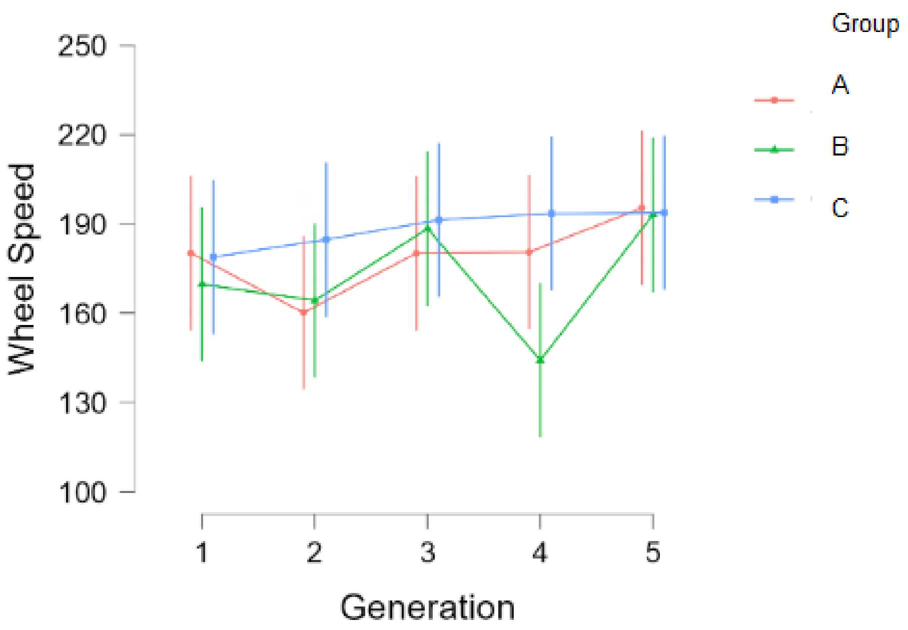
“It could be that one needs a much longer track for a configuration like this to be good. So you can gain some speed, since it is quite slow in the beginning”, Group B, Chain 2, Participant 5.

Appendix B

Figures and Tables

Figure B1.

Model 2 - The Average Wheel Speed between Treatments (groups), described i Results 5.1



Note. Shows the average of wheel speeds (all five trials for all participants) across generations, divided by treatments. Wheel speed was measured in m/h. Error bars: 95% *(CI).*

Table B1.

Estimated Marginal Means and Standard Error for Model 1, described in Results 5.1.

Model 1				
Generation	**EMM of Wheel Speed across Generations**	**Standard Error**	**95% (CI) : Lower**	**95% (CI) : Upper**
1	155.353	5.440	144.691	166.016
2	166.453	5.861	154.965	177.941
3	174.066	6.140	162.032	186.099
4	167.507	6.538	154.694	180.321
5	183.410	5.683	172.271	194.549

Table B2.

Estimated Marginal Means and Standard Error for Model 2, described in Results 5.1.

Model 2				
Group	**EMM of Wheel Speed between Treatments**	**Standard Error**	**95% (CI) : Lower**	**95% (CI) : Upper**
A	171.514	5.268	161.189	181.838
B	168.323	5.268	157.999	178.648
C	168.237	5.268	157.912	178.561

Table B3.

Estimated Marginal Means and Standard Error for Model 3, described in Results 5.2.

Model 3				
Group	**EMM of Transferred Trials between Treatments**	**Standard Error**	**95% (CI) : Lower**	**95% (CI) : Upper**
A	179.331	6.386	166.814	191.848
B	171.995	6.386	159.478	184.512
C	188.514	6.386	175.997	201.031

Table B4.

Estimated Marginal Means and Standard Error for Model 4, described in Results 5.3.

Model 4				
Generation	**EMM of Test Scores across Generations**	**Standard Error**	**95% (CI) : Lower**	**95% (CI) : Upper**
1	5.944	0.421	5.118	6.770
2	6.667	0.421	5.841	7.493
3	6.500	0.421	5.674	7.326
4	6.611	0.421	5.785	7.437
5	7.000	0.421	6.174	7.826

Table B5.

Estimated Marginal Means and Standard Error for Model 5, described in Results 5.3.

Model 5				
Group	**EMM of Test Scores between Treatments**	**Standard Error**	**95% (CI) : Lower**	**95% (CI) : Upper**
A	6.800	0.381	6.054	7.546
B	6.367	0.381	5.620	7.113
C	6.467	0.381	5.720	7.213

Table B6.

Estimated Marginal Means and Standard Error for Model 6, described in Results 5.4.

Model 6				
Generation	**EMM of Graded Theory Transmissions across Generations**	**Standard Error**	**95% (CI) : Lower**	**95% (CI) : Upper**
1	0.889	0.278	0.345	1.433
2	1.333	0.278	0.789	1.877
3	1.778	0.278	1.234	2.322
4	2.056	0.278	1.512	2.599
5	1.667	0.278	1.123	2.211

Table B7.

Estimated Marginal Means and Standard Error for Model 7, described in Results 5.4.

Model 7				
Group	**Estimated Marginal Mean of Graded Theory Transmissions between Treatments**	**Standard Error**	**95% (CI) : Lower**	**95% (CI) : Upper**
A	2.100	0.252	1.606	2.594
B	1.333	0.252	0.840	1.827
C	1.200	0.252	0.706	1.694

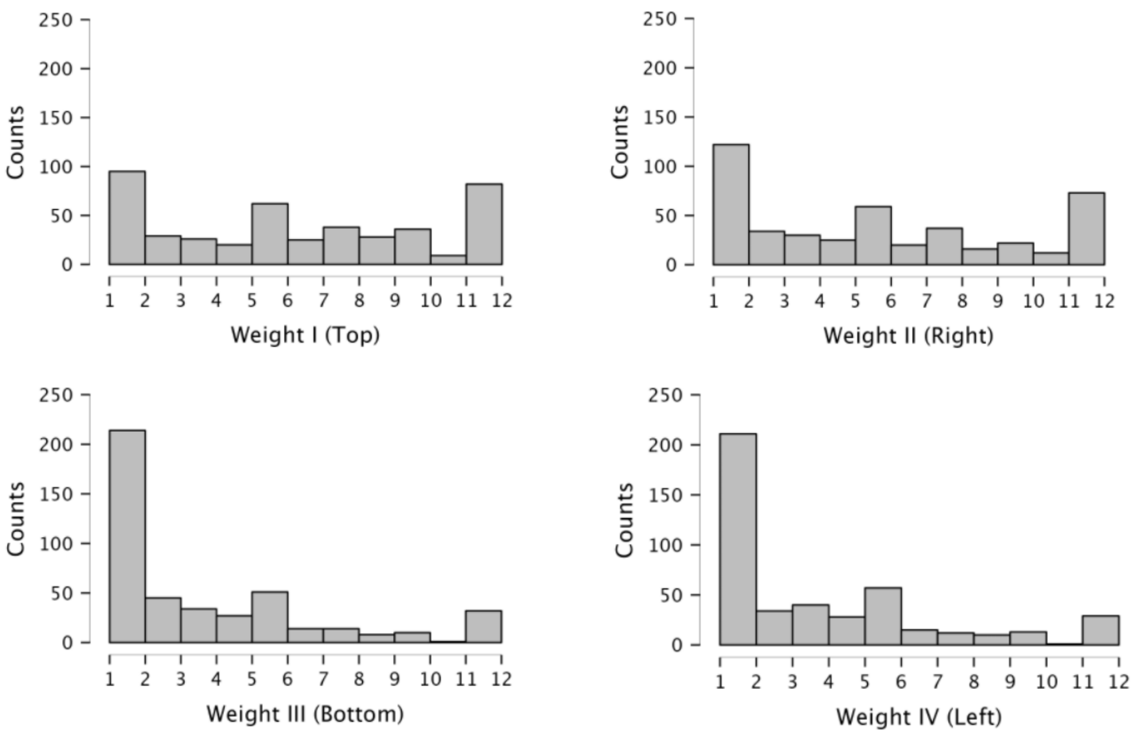
Table B8.

Mean of Weight Placements along each spoke

	Weight I (Top)	Weight II (Right)	Weight III (Bottom)	Weight IV (Left)
Mean	6.578	5.933	3.878	3.953
Standard Deviation	3.816	3.874	3.334	3.310
Minimum	1.000	1.000	1.000	1.000
Maximum	12.000	12.000	12.000	12.000

Figure B2

Weight placements along each spoke



Note. Depicts the frequencies of placement. Each bar graph refers to a seperate spoke.

Figure B3

Archeulean Hand-Axe

Note. Archeulean hand-axe, created by the teacher in the experiment by Putt et al. (2014).

Appendix C

Construction of the equipment

C1 - Construction of the apparatus

The apparatus used in this study is a replication of the design of the apparatus from previous studies by Derex et al. (2019) and Osiurak et al. (2021) in as much detail as the situation allowed. All quantitative measures in terms of weights and distances were obtained. The two main components of the experiment apparatus are the wheel and the rail structure which the wheel spins along (see Figure 1). A pre-pilot study was conducted to identify obstructions in the design, implementation of instructions and usage of interface. The pre-pilot study included five participants who were novice to the goal of their involvement. Data from the pre-pilot study was not included in the official collection of data and was solely used to optimise the conduction of the main study. Errors found in the experimental design used in the pre-pilot were taken into account and edited for the main study. The main version is described below.

C.1.2 - Building of the wheel

The wheel with its four spokes and the centre axis were held together with a three axle tube clamp (Appendix C - Figures C1 & C2). The spokes as well as the axis were made out of a wooden round bar with 28 mm in diameter. The spokes extended 410 mm from the centre of the axis and were marked up in 28 mm discrete segments with red heat shrink tubing. The centre axis, fitted perpendicular to the spokes through the three axle tube clamp, extended differently on either side of the tube clamp to compensate for the asymmetric design of the tube clamp. The length of the centre axis wooden bar was the sum of the distance through two 500 g weight discs, two barbell clamps, one 11 mm spacer and one three axle tube clamp. An 8 mm longitudinal hole through the centre axis was machined on lathe to obtain a centred straight hole. An 11 mm thick plywood spacer was cut out and machined on lathe and mounted around the centre axis on the side of the tube clamp where the centre axis extended longer. The purpose of the spacer was to keep the 500 gr weight discs on equal distance from the centre of the spokes, which was critical to maintaining a straight path along the rail track. A M8 threaded steel rod was snuggly fitted through the hole in the centre axis and on either side 40 mm rubber tube and two pairs of nuts to lock all parts in place. The purpose of the rubber tube was to prevent sliding against the rails, while the locking nuts had the purpose to prevent the rubber tube from twisting around the threaded steel rod.

Four pieces of adjustable weights were composed out of barbell clamps, multiple metal washers held in place with M5 bolts and nylock nuts. Each unit weighed 100 g ± 0.5 g, controlled on a two decimal digital scale, and placed along each spoke. The weight units were composed of the same type of barbell clamp that were used to lock the 500 g weight disc on the centre accel in position.

To battle a, wider than tolerable, dimensional range in the accuracies of many of the individual parts the wheel were fixated in a jig (see Appendix C - Figure C1) to maintain perpendicularity of the spokes and centre accel. The spokes and accel were glued together inside the three axle clamp with expanding polyurethane glue. The specific glue was picked for its ability to glue metal together with wood as well as for filling out the excess space in the tube clamp making a gapless bond able withstand the forces involved in the experiment. One design problem found during the course of the study was that the rubber tube on the centre axis got unevenly squeezed lengthwise, thereby radially expanded, causing a difference in radius on either side and imposing a slight deviation from the wheel spinning straight on the rails. A similar twisting problem occurred between the weight discs and the wooden centre axis. During the course of the experiment the problematic sectional parts mentioned above were continuously replaced to avoid confounding performance of the apparatus.

C.1.3 - Building of the rails

The rail structure, along which the wheel spun, was welded and bolted together out of different types of aluminium profiles. The rails were constructed of two separate 2000 mm long aluminium L profiles, positioned in a 14 degree angled slope. Two stop blocks made out of triangular shaped low friction plastic were bolted in the lower end of the rails. The aim of the shape whas to let the wheel slidingly climb a few centimetres and gradually decrease rotation speed. The frame structure holding the rails were made 40 mm wider than the width of the rails to give space in case of further adjustments for the wheel to spin free of sideways friction from the rails. A counter balanced lever was placed in the highest part of the frame structure right above the start position of the wheel. The lever held the wheel in the desired starting position, perpendicular to a horizontal line, enabling the wheel to maintain an equal starting position for each start. In each four ends of the two base beams 8 mm vertical holes were drilled to hold bolts providing adjustable feet to obtain levelled support for the rail.

C2 - Interface and time measurement

The design of the interface was copied with modifications from Derex et al (2019). A web interface together with the timing was made up of three separate parts that work together (see Appendix C - Figure C3 for diagram). Firstly, an ESP32 microcontroller for the actual timekeeping. The microcontroller ran a code written in Arduino Programming Language and was connected to the start and stop switches on the apparatus' rail. Two time switches placed at the start and at the 1 m point of the track forwarded events to the server when the voltage in the switches changed. Secondly, a server written in Node.js that handled communication with the ESP32. The server stored data in a csv file from which the web interface could download the data through a REST API. The server also forwarded the time measurements from ESP32 to the interface through web sockets (socket.io). Thirdly, a web interface written in Javascript built with React that handled the regulation and display of data available to participants. The interface retrieved data from the server through a REST API and updated new data through the same API. A web socket connection was established, for each trial through validation, that listened for a triggering event containing new data output. See Figure 1 for visual reference to the full set up and Figure 2 for visual reference to the interface.

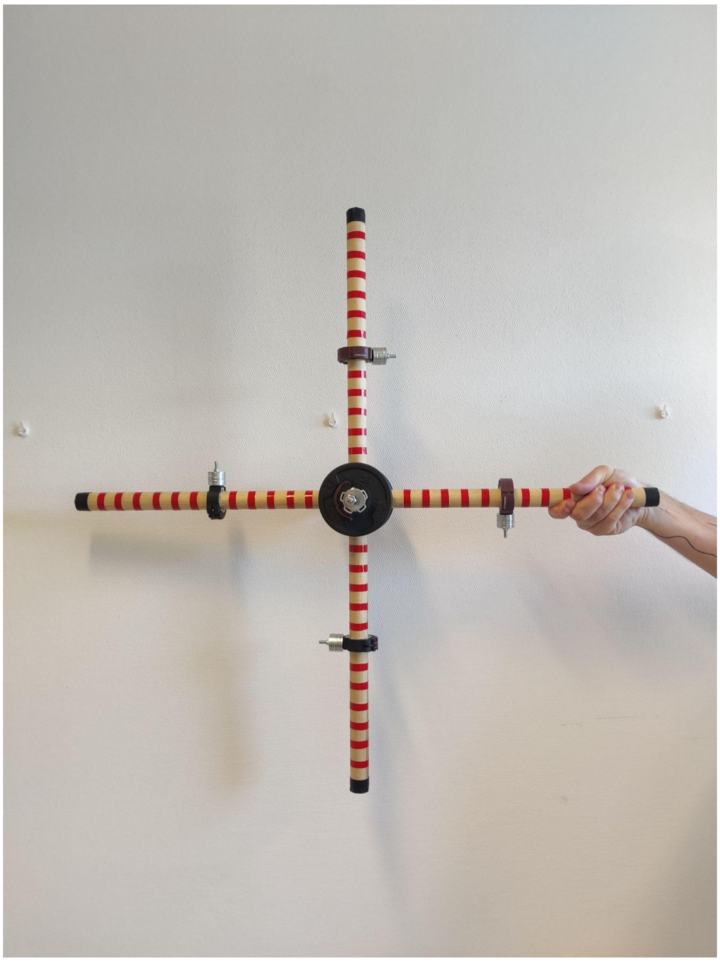
Figure C1

Fixation of the wheel

Note. The wheel was fixated in a jig in the creation process (see Appendix C1.2 for description)

Figure C2

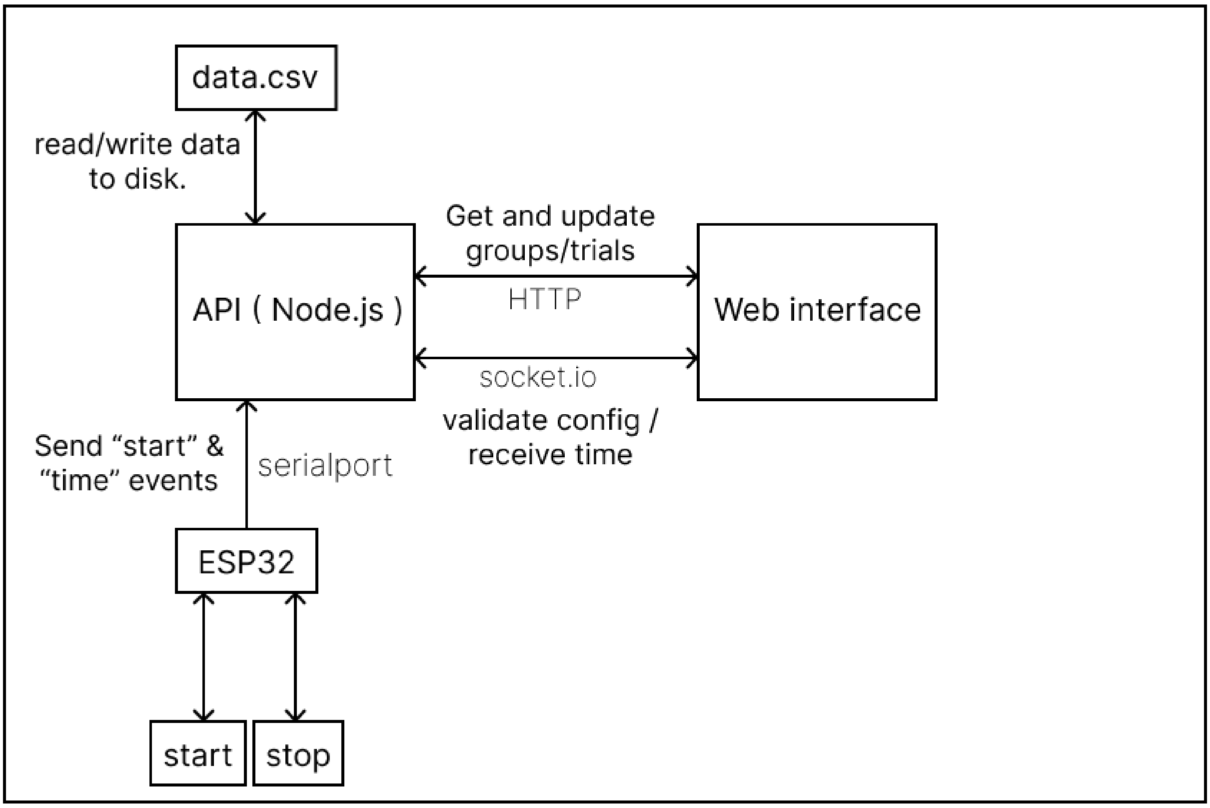
The wheel



Note. The wheel that is used throughout the study. Each of the red stripes marks a discrete position on which a weight could be placed. The weights here are placed at position 6 on each of the four spokes, which was also the initial position before each participant was presented with the option of alteration.

Figure C3

Flow Scheme; Time keeping and interface

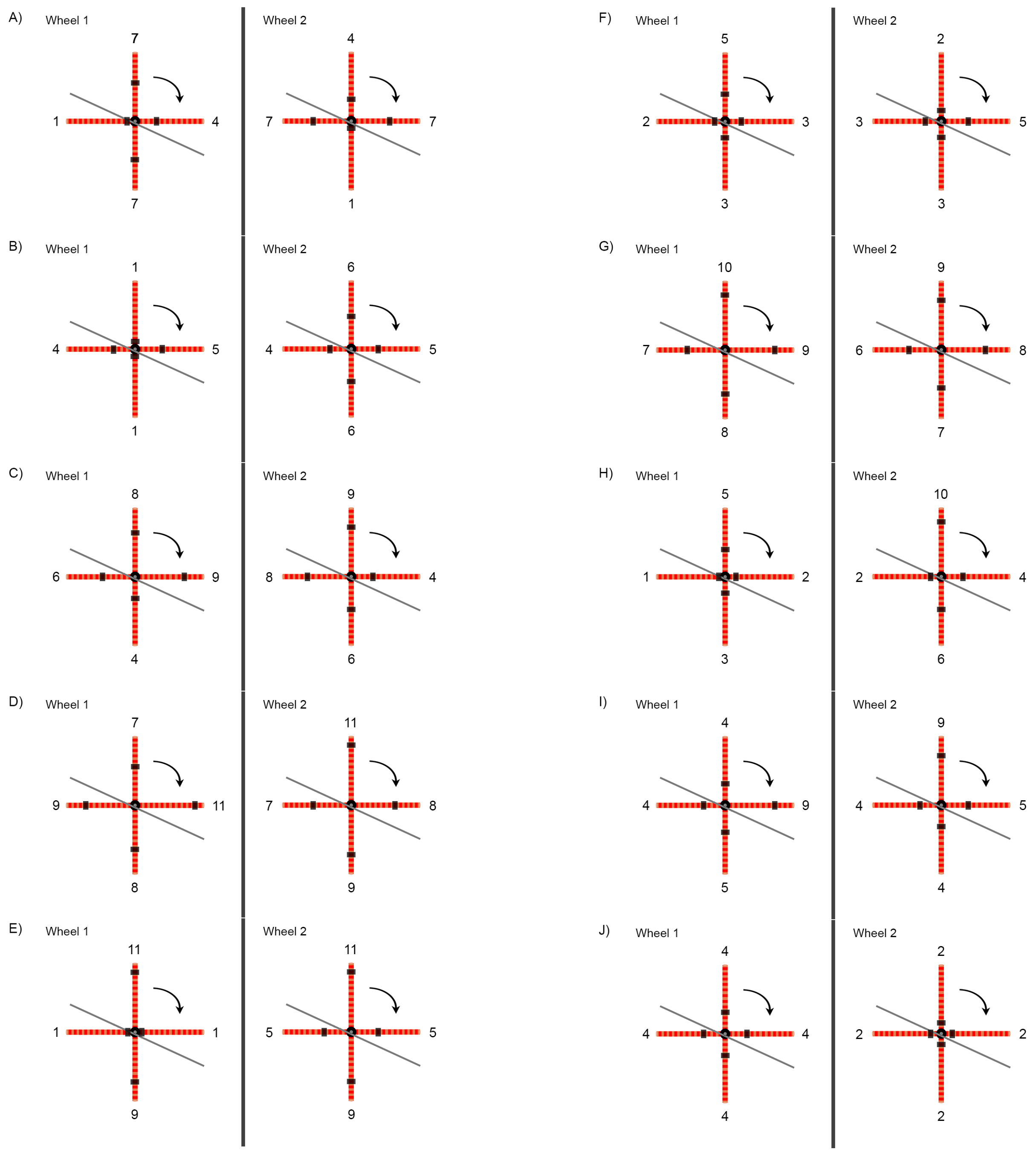


Appendix D

Knowledge test and instructions

Figure D1

Knowledge test



Note. The complete knowledge test, each wheel pair presented separately for participants. Wheel pairs A, C, D, F and I differed in centre of mass distribution. Wheel pairs B, E, G, H and J differed in their moment of inertia. Correct answers: A (2), B (1), C (1), D (2), E (1), F (1), G (2), H (1), I (2), J (2).

D1 - Instructions

Instructions differed slightly with regards to assigned treatment and chain position. The following instructions is the instructions handed to participants in group A and B, with the chain position 2-4, in Swedish:

Framför dig finns ett hjul och två skenor. Hjulet har 4 ekrar och en vikt kan flyttas längs varje eker. Din uppgift är att placera vikterna för att få hjulet att rulla ned så snabbt som möjligt längst den markerade banan. Varje vikt kan placeras på en av 12 distinkta positioner längst varje eker. Position 1 är närmast axeln. Position 12 är längst bort från axeln. Konfigurationer, alltså vikternas placering, väljs via datorn och ställs sedan in med hjälp av funktionären.

När du väl har valt din konfiguration kommer försöksledaren att flytta vikterna på det fysiska hjulet. Du kommer att genomföra 5 försök. Efter varje försök kommer hjulets rull-tid att visas på skärmen. Innan varje försök kommer du ha möjligheten att se dina två senaste konfigurationer och dess prestanda på skärmen genom att klicka på de markerade siffrorna högst upp på skärmen. Detta kan göras när som helst under experimentet.

Utvecklingen av hjulet är en kollektiv uppgift. Även om du är ensam i rummet är du en del av en längre kedja av deltagare. De två sista försöken från varje deltagare kommer att överföras till nästa deltagare i kedjan. Denna information kan hjälpa deltagarna att uppnå bättre prestationer. Innan varje försök har du möjligheten att se föregående deltagares två sista konfigurationer. Registret kan nås under hela experimentet och låter dig se föregående deltagares två sista konfigurationer och dess prestanda på skärmen genom att klicka på de markerade siffrorna högst upp på skärmen.

Tillsammans med informationen om konfigurationer och prestanda, kommer du också att få en kort beskrivning av föregående deltagares tankar om hur vikterna kan placeras på hjulet. Beskrivningen kommer att vara tillgänglig under alla dina försök. Efter dina fem försök kommer du att bli ombedd att på samma sätt skriva ner dina tankar om vad som påverkar hjulet. Din förklaring kommer att skickas till nästkommande deltagare.

Du är nu redo att börja. Efter varje försök kommer prestationen att visas i tid på skärmen. Var försiktig, vissa konfigurationer hindrar hjulet från att nå slutet av banan.

D2 - Summary

Summaries differed slightly with regards to assigned treatment and chain position. The following summary is the summary handed to participants in group A and B, with the chain position 2-4, in Swedish:

- *Ditt mål är att få hjulet att rulla så fort som möjligt. För detta har du fem försök. Dina två sista försök kommer att visas för nästa deltagare.*

- *Varje eker har 12 positioner. Du flyttar vikterna på datorn, funktionärer flyttar vikterna på hjulet.*

- *Du har möjlighet att se resultatet och konfigurationen från dina två senaste försök. Detta görs genom att klicka på de markerade siffrorna högst upp på skärmen. Är du till exempel på försök tre kan du se dina två första försök. När du är på försök fyra kan du se försök två och tre, osv.*

- *Du har även möjlighet att se föregående deltagares två sista försök. Detta görs genom att klicka på de markerade siffrorna högst upp på skärmen. Detta kan du komma åt under alla dina försök.*

- *Under alla dina försök kommer du att kunna se föregående deltagares tankar kring hjulet och placeringen av vikterna.*

- *Efter dina fem försök kommer du att skriva en kort beskrivning av dina tankar kring hjulet och placeringen av vikterna.*