Michał Szczepanik

Neural Correlates of Emotional Contagion in Humans: Familiarity Between Participants Does Not Enhance Transmission of Fear

Phd thesis
Completed in the Laboratory of Brain Imaging
of the Nencki Institute of Experimental Biology
Polish Academy of Sciences

SUPERVISOR:
Ewelina Knapska, Ph.D., D.Sc.
AUXILLARY SUPERVISOR:
Marek Wypych, Ph.D.

Warsaw, 2021

Fear is contagious. You can catch it. Sometimes all it takes is for someone to say that they're scared for the fear to become real. Mo was terrified, and now Nick was too.

Neil Gaiman, The Graveyard Book

Abstract

The ability to learn through observation of others is advantageous. When gathering information about threats, observational learning is favourable because direct experiences can be costly. Like other social species, humans are well equipped to perceive and interpret cues in the behaviour of others, and the process involves the perception of emotions and emotional contagion.

This thesis describes two experiments, which used observational fear conditioning to investigate the social transfer of emotions in pairs of people (demonstrator and observer). An established experimental paradigm was modified to increase naturalness and ecological validity by employing live observation. The demonstrator underwent a typical differential conditioning task, in which a neutral stimulus was paired with an aversive electrical stimulation. The observer witnessed this situation without experiencing the stimulation. Next, the observer was confronted with the same task, but the aversive stimulation did not occur. Electrophysiological (skin conductance, acoustic startle) and neuroimaging (functional magnetic resonance imaging) methods were used to measure the process and outcome of observational conditioning.

The main goal of the doctoral thesis was to investigate whether psychophysiological reactions and patterns of brain activity during observational fear conditioning differ depending on the familiarity between demonstrator and observer. Because familiarity is implicit in friendship, friend dyads were studied, and learning from friends was compared to learning from strangers in a group design. According to a commonly accepted theoretical model of Preston and de Waal, it was hypothesised that potential differences in observational fear conditioning between familiar and unfamiliar dyads could be related to a different level of empathy shown by the two groups.

Results suggest that physiological outcomes of observational fear conditioning are dependent on declarative learning. Observed patterns of brain activation confirm joint involvement of networks engaged in fear and social perception. The anterior insula and posterior superior temporal sulcus are suggested as crucial hubs of these networks, and the engagement of the amygdala, anterior cingulate cortex, and fusiform face area are highlighted. Finally, no differences were found between observation of friend and stranger, leading to a conclusion that familiarity between participants does not influence observational fear conditioning. It can be assumed that in the absence of other differentiating factors, friendship alone does not enhance learning, and strangers are as reliable sources of threat information as friends.

Streszczenie

Zdolność uczenia się poprzez obserwację innych jest korzystna. Przy pozyskiwaniu informacji o zagrożeniach, możliwość uczenia obserwacyjnego jest preferowana, ponieważ bezpośrednie doświadczenia mogą być kosztowne. Podobnie jak inne gatunki społeczne, ludzie są dobrze przygotowani do postrzegania i interpretowania wskazówek w zachowaniu innych, a proces ten obejmuje percepcję emocji i zarażanie emocjonalne.

Niniejsza praca opisuje dwa eksperymenty dotyczące transferu emocji między parami ludzi (demonstrator i obserwator). Istniejący protokół obserwacyjnego warunkowania strachu został zmodyfikowany w celu zwiększenia naturalności i trafności ekologicznej poprzez zastosowanie obserwacji na żywo. Demonstrator wykonywał typowe zadanie warunkowania różnicowego, w którym neutralny bodziec był łączony z awersyjną stymulacją elektryczną. Obserwator był świadkiem tej sytuacji, nie doświadczając stymulacji. Następnie obserwator został skonfrontowany z tym samym zadaniem, ale awersyjna stymulacja nie wystąpiła. W badaniach wykorzystano pomiar przewodnictwa skóry i odruchu wzdrygnięcia oraz funkcjonalny rezonans magnetyczny.

Głównym celem pracy doktorskiej było zbadanie, czy reakcje psychofizjologiczne i wzorce aktywności mózgu podczas obserwacyjnego warunkowania strachu różnią się w zależności od stopnia znajomości między demonstratorem i obserwatorem. Ponieważ przyjaźń jest przykładem bliskiej znajomości, eksperymenty przeprowadzono z udziałem par przyjaciół, a uczenie się od przyjaciół porównano z uczeniem się od nieznajomych. Zgodnie z powszechnie akceptowanym modelem teoretycznym Preston i de Waala, postawiono hipotezę, że potencjalne różnice w obserwacyjnym warunkowaniu strachu pomiędzy znajomymi i nieznajomymi mogą być związane z różnym poziomem empatii wykazywanym przez obie grupy.

Wyniki sugerują, że fizjologiczne miary obserwacyjnego warunkowania strachu są zależne od deklaratywnego uczenia się zależności między bodźcami. Zaobserwowane wzory aktywacji mózgu potwierdzają wspólne zaangażowanie sieci zaangażowanych w strach i percepcję społeczną. Wskazują one, że przednia część wyspy i tylna część bruzdy skroniowej górnej są kluczowymi węzłami tych sieci; ważne jest też z zaangażowanie ciała migdałowatego, przedniej części kory zakrętu obręczy i zakrętu wrzecionowatego. W analizach nie stwierdzono różnic między obserwacją przyjaciela i nieznajomego, co prowadzi do wniosku, że znajomość między uczestnikami nie wpływa na obserwacyjne warunkowanie strachu. Można założyć, że przy braku innych czynników różnicujących, sama przyjaźń nie wzmacnia uczenia się, a osoby obce są tak samo wiarygodnym źródłem informacji o zagrożeniu jak przyjaciele.

Acknowledgements

A doctorate is a journey that leaves one profoundly changed compared to who one was initially. This doctorate was no different. Throughout the time I spent at the Nencki Institute, several people influenced me and, more or less directly, this thesis.

First of all, I would like to thank Anna Kaźmierowska, with whom I worked closely in the same research project as a fellow doctoral student. Without her enthusiasm, positive attitude and psychological knowledge, the path to completion of the project would have been much more difficult.

I was fortunate to have Ewelina Knapska as my supervisor and, without a doubt, mentor. She has broad psychological and biological knowledge and a unique perspective bridging animal and human studies. Whenever I encountered problems around my research, I always felt that she was available to discuss them, and her can-do attitude and optimism were genuinely contagious. I am very grateful for all our conversations on both scientific and meta-scientific subjects. Her support exceeded this research project, and her advice helped me choose my post-doctoral path with confidence.

My co-supervisor, Marek Wypych, was always happy to talk about science and life in general. He taught me a lot of technical details about fMRI. He shared with me his honest outlook on the scientific career, and wanted me to make informed decisions. His unbridled critical and polemical passion helped make multiple sections of this thesis more precise. On a personal note, we shared a passion for trekking, cheese, and gadgets. He inspired me to run more regularly, which gave me a vital way to clear my mind after work.

Jarosław Michałowski was another essential contributor to the project, who brought expertise in clinical psychology and the psychophysiology of fear. Andreas Olsson, an expert in observational fear conditioning, stepped in with invaluable advice at critical moments. We mostly met online (with Jarosław located in Wrocław and Andreas in Stockholm), but they were always involved.

Artur Marchewka was the person who introduced me to fMRI already during my undergraduate studies. His knowledge and passion enticed me to start this doctorate. Throughout this research project, I could rely on his advice, especially regarding the neuroimaging study. As the head of the Laboratory of Brain Imaging, he cared a lot about the work-life balance, was willing to leave me the freedom to pursue side projects, and was direct in his feedback. Moreover, he created a collaborative environment that promoted its members' well-being, destigmatised failure, and celebrated success.

Consequently, the Laboratory has grown to be a group of enthusiastic people with

different backgrounds, and I feel privileged to have been a part of it. Among these people, I would especially like to thank Małgorzata Wierzba. She was the archetypal senior colleague: she was a keen listener whenever I needed to talk about a problem and often gave thoughtful advice. Bartosz Kossowski shared my interest in the technical side of research, taught me a lot about MRI sequences, and partnered on several side projects. On the few occasions when we programmed together, his tendency to move fast and break things complemented my tendency to overthink. Last but not least, Dawid Droździel oversaw the research equipment and ensured that all experiments could run smoothly. He helped during several data collection sessions and was a great companion when we shared the same office.

This work would not have been possible without institutional support. The research project was funded by the National Science Centre grant 2015/19/B/HS6/02209, awarded to Ewelina Knapska. Additionally, I have been supported by a doctoral scholarship from the Nencki Institute.

Abbreviations

ACC . . . . . Anterior cingulate cortex
AI . . . . . Anterior insula
ANOVA . . . . . Analysis of variance
BES-A . . . . . Basic Empathy Scale for Adults (questionnaire)
BF . . . . . Bayes factor
BOLD . . . . . Blood oxygenation level dependent (signal)
CR . . . . . Conditioned response
CS . . . . . Conditioned stimulus
DVARS . . . . . Derivative of root mean square variance over voxels (VARS)
EDA . . . . . Electrodermal activity
EEG . . . . . Electroencephalography
EMG . . . . . Electromyography
EPI . . . . . Echo-planar imaging
FD . . . . . Framewise displacement
FFA . . . . . Fusiform face area
FIR . . . . . Finite impulse response
FPS . . . . . Fear potentiated startle
FWE . . . . . Familywise error
FWER . . . . . Familywise error rate
GLM . . . . . General linear model
GSR . . . . . Galvanic skin response
HR . . . . . Heart rate
HRF . . . . . Haemodynamic response function
ICBM . . . . . International Consortium of Brain Mapping
IQR . . . . . Interquartile range
ITI . . . . . Intertrial interval

MCC . . . . . Mid-cingulate cortex
MFQ . . . . . McGill Friendship Questionnaire
MNI . . . . . Montreal Neurological Institute
MR . . . . . Magnetic resonance
MRI . . . . . Magnetic resonance imaging
PAM . . . . . Perception-action model
PPI . . . . . Psychophysiological interaction
ROI . . . . . Region of interest
SCL . . . . . Skin conductance level
SCR . . . . . Skin conductance response
SD . . . . . Standard deviation
SDI . . . . . Serial digital interface
SEM . . . . . Standard error of the mean
STAI . . . . . State and Trait Anxiety Index
STS . . . . . Superior temporal sulcus
SVC . . . . . Small volume correction
SWE . . . . . Skala Wrażliwości Empatycznej (Empathic Sensitiveness Scale)
TPJ . . . . . Temporoparietal junction
US . . . . . Unconditioned stimulus
aMCC . . . . . Anterior mid-cingulate cortex
dACC . . . . . Dorsal anterior cingulate cortex
fMRI . . . . . Functional magnetic resonance imaging
pSTS . . . . . Posterior superior temporal sulcus
rTPJ . . . . . Right temporoparietal junction
rpSTS . . . . . Right posterior superior temporal sulcus
vACC . . . . . Ventral anterior cingulate cortex

Contents

Chapter 1

Introduction

1.1 The legacy of I. Pavlov

The general concept of conditioning entered scientific discourse (and the minds of the general public) due to the works of Ivan Pavlov at the turn of the 20th century (Cambiaghi & Sacchetti, 2015; Pavlov, 1928). Although Pavlov's research stemmed from an investigation of the digestive process and concentrated primarily on salivary reflex, the notion of conditioning was quickly expanded into different areas of behaviour, of which fear conditioning is of particular interest for the current work.

Early examples of fear conditioning in humans can be traced to the infamous "Little Albert" experiment (Watson & Rayner, 1920). Its authors claimed to have elicited a conditioned fear response to a white rat by presenting the animal to an 11 months old child while simultaneously striking a steel bar to produce a loud, aversive sound. The response was allegedly generalised to other animals and particular objects resembling the rat, and the conditioned fear persisted over time. Although the experiment's scientific and ethical merits are highly doubtful by modern standards, its description shows that the idea of fear conditioning, and its relation to the development of phobias, has been present in the psychological literature already for a century.

Furthermore, studies on vicarious, or social, fear conditioning (a process of conditioning through observation of others rather than through one's own experiences) can be traced at least to the 1960s (Bandura & Rosenthal, 1966; Berger, 1962). For example, in Bandura and Rosenthal's experiment, participants who watched another person exhibiting convincingly feigned pain reactions preceded by a buzzer sound during an unrelated motor task exhibited conditioned skin conductance responses without experiencing painful stimulation themselves. The strength of conditioned responses was modulated by the observer's physiological arousal, which was manipulated by verbal instructions and epinephrine injections, and was strongest in the group with moderate arousal level.

Recently, a formal protocol of observational fear conditioning has been proposed (Haaker, Golkar et al., 2017). It uses video recordings of an actor experiencing an aversive electrical stimulation paired with a visual cue. The recordings are used as social stimuli

which are viewed by a subject. The protocol provides clear guidance on conducting the experiments while also leaving room for adaptations to specific research questions and types of measurements.

The examples above demonstrate that fear conditioning has been a subject of scientific investigation across the last century. It is also a key concept for the experiments described in this thesis, in which I used a modified version of the observational fear conditioning protocol to investigate how friendship affects transmission of fear. The following sections of the Introduction will further introduce several aspects of fear conditioning, selected methodology and research questions.

1.2 What is Pavlovian fear conditioning?

Fanselow and Wassum (2015) suggested that Pavlovian conditioning should be defined as "the process whereby experience with a conditional relationship between stimuli bestows these stimuli with the ability to promote adaptive behaviour patterns that did not occur before the experience". This is a general definition, emphasising that conditioning is a process of learning about dependent relationships between stimuli. The process may involve different kinds of stimulus relationships: for instance, some stimuli may predict rewards (appetitive conditioning), while others may inform about danger (fear conditioning, also named threat conditioning). In a narrower sense, fear conditioning is a "laboratory model for learning to predict threat through association of initially neutral stimuli (conditioned stimuli, CS) with aversive outcomes (unconditioned stimuli, US)" (Ojala & Bach, 2020). As an outcome of conditioning, the CS alone begins to elicit anticipatory conditioned responses (CR). The CR can manifest, for instance, by autonomic activation[1], and observation of the CR is taken as evidence that the association between CS and US has been formed (Ojala & Bach, 2020).

A large family of experimental protocols with shared principles and methodology (Lonsdorf et al., 2017) is related to fear conditioning. Two notions are crucial for the remainder of the thesis. First, a typical fear conditioning experiment is likely to be divided into two stages, acquisition and test. As the names suggest, acquisition involves paired presentations of the CS and US so that an association may be formed, while in the test stage, the CS are presented alone to evaluate the existence (and persistence) of the CR. Second, especially in humans, a differential conditioning task is commonly employed. In this task, there are two CS, one of which (CS+) is paired with an aversive US, while the other (CS–) is not. With such a design, the CS– provides a within-subject baseline condition, to which the reactions elicited by CS+ can be compared.

[1] Going back to the general definition of Pavlovian conditioning, "a CR is any response that can be directly attributed to that conditional relationship" and, notably, does not have to be identical with the reaction to the US (Fanselow & Wassum, 2015).

1.3 Social learning of fear

Being able to learn about threats present in the environment is of crucial importance for many animals, including humans. However, learning from direct experience can be costly; examples can range from a wild animal being attacked by a predator when feeding in an exposed location to a home cook getting burnt when touching a scorching pot lid when cooking. This is why being able to learn through observation of others is equally, if not more, important. To build on previous examples, an animal may learn to avoid a given place after seeing their conspecific narrowly escaping an attack, and a cook can learn always to wear gloves when removing lids after seeing their friend get burned.

Such vicarious learning is driven by perceiving emotional responses of others, which can be sufficient for the transfer of fear even if actual harm is not being done - for example, fearful reactions of adult monkeys in the mere presence of a snake can act as powerful cues for their younger conspecifics (Mineka et al., 1984). Although the examples given so far are adaptive, it is worth noting that social learning can at times be also maladaptive. Examples of maladaptive social learning include the development of childhood fears (Askew & Field, 2007) or, in a broader context, the development of anxiety as a result of media exposure to disasters and large-scale violence (Hopwood & Schutte, 2017).

Learning in the examples above is based on forming associations between events. Considering that Pavlovian threat conditioning is one of the primary laboratory models of associative learning (Fanselow & Wassum, 2015), it is not surprising that it also became a basis for social learning studies. The transition to a vicarious conditioning protocol is achieved by introducing two participants: a demonstrator (also referred to as a learning model), whose expressions of discomfort (or distress) serve as social cues, and an observer, who learns about the threat through watching the demonstrator, without experiencing the source of discomfort first-hand. Given the social nature of multiple species and the significance of social learning, it is reasonable to assume that observational acquisition of fear is underlain by specific neural mechanisms (Adolphs, 2009), as well as those known from direct conditioning (Olsson et al., 2007). The specific questions related to the neural architecture of fear can naturally be addressed in more detail in animal studies, but human research can also provide significant insight.

1.4 Fear conditioning across species

Fear conditioning has been studied extensively in humans and non-human animals, thus allowing cross-species comparisons and broad insight into its biological mechanisms. This is possible by employing analogous experimental protocols (similar both conceptually and methodologically), which often use similar outcome measures. A detailed overview of the cross-species methodology, highlighting key similarities and notable differences, can be found in (Haaker et al., 2019) and is summarised below.

In general, a procedure involving a CS which predicts an aversive US and reliably leads to forming an association between the two can be easily applied in both kinds of

studies. For the CS, visual and auditory modalities are most commonly used. They can be applied across species; however, the former is typically used in humans and the latter in rodent experiments. For the US, electrical stimulation is used most commonly, but other stimuli, such as loud white noise bursts or air-blasts[2] can also be applied across species. Notably, though, in human studies, the intensity of electrical stimulation can be adjusted by the participant according to the perceived level of unpleasantness and is arguably less aversive than in the case of foot shocks applied to rodents. The timing of stimulus presentation within a session can be similar across species; however, multi-day paradigms with several sessions are much more common in rodent studies. Additionally, most rodent studies use a single CS. In contrast, human studies are usually based on differential conditioning, where one CS is paired with the aversive stimulus while the other is not.

One popular physiological measure of fear conditioning outcome which can be applied similarly in human and rodent studies is fear-potentiated startle. The acoustic startle response is a primitive defensive reaction, which can be elicited by a short burst of white noise. It is mediated by a relatively simple neuronal circuit involving the brainstem and centromedial amygdala, conserved across species (Koch, 1999; Kuhn et al., 2020). In rats, it can be measured through a whole-body reaction as a pressure exerted on the ground. In humans, it is based on the eyeblink measured with facial electromyography of the orbicularis oculi muscle.

The above considerations regarding stimulus presentation and outcome measures were formulated concerning non-social fear conditioning, but they are applicable just as well for observational protocols. Here, several notable vicarious fear conditioning studies in non-human animals can serve as an inspiration and reference points for human studies. In a seminal study, (Mineka et al., 1984) demonstrated how young cage-reared rhesus monkeys acquire fear of snakes through observation of their wild-reared parent's reactions. Rodent models have also been studied in this context. Both mice (Jeon et al., 2010; Jeon & Shin, 2011) and rats (Andraka et al., 2021; Atsak et al., 2011; Knapska et al., 2006) can either interact with their previously conditioned cagemates or observe them during a fear conditioning procedure involving foot shocks. Such protocols allow for observation of threat-specific behavioural and physiological responses in the observers. Moreover, the observation protocols are conceptually similar to the observational fear conditioning protocol used in the current thesis.

1.5 Emotional contagion, empathy, and socially acquired fear

Given its inherently social nature, vicarious fear conditioning can be situated in the realms of emotional contagion and, in broader terms, empathy. Emotional contagion is defined as "emotional state matching between a target and an observer" (de Waal & Preston, 2017). Empathy can be understood as "the ability to respond to and experientially

[2] An interesting choice for human experiments is to use a shrieking scream. For example, one study (Hamm et al., 1989) used a sample from a 1977 song "Frankie Teardrop" by Suicide, which I have to admit can be very effective.

share the feelings of others, which eventually leads to a better understanding of their inner emotional and mental states" (Adriaense et al., 2020). It is thought that the term was coined by Titchener (1909), who proposed it as a corresponding term for german *Einfühlung*, best explained as "feeling for"[3]. That being said, defining such a complex phenomenon is not easy. In a review of human empathy literature, Cuff et al. (2016) were able to identify 43 distinct definitions of empathy, which they broke down into eight main themes[4] which the definitions attempt to reflect.

One of the most influential theoretical framings of empathy has been proposed by Preston and de Waal (2002), who consider it a phylogenetically continuous phenomenon, exhibited by humans and other animal species, and thus requiring a broad definition. The authors proposed that at its core, empathy is based on the perception-action model (PAM), summarised as follows: "attended perception of the object's state automatically activates the subject's representations of the state, situation, and object, and that activation of these representations automatically primes or generates the associated autonomic and somatic responses, unless inhibited" (Preston & de Waal, 2002). Several notions are crucial for this definition. First, PAM stems from the perception-action theory of motor behaviour. Second, the word 'representation' is used here in the context typically employed by cognitive science to broadly describe neuronal connections in the brain which encode information. Third (as the authors point out in response to peer commentary published together with the original article), although said representations are activated automatically by associated perceptions, they are not uncontrollable. Instead, they should be gated by attention and remain susceptible to other mediating factors.

The PAM provides a potential underlying mechanism for explaining the modulation of empathy by familiarity or similarity. Since representations can change with experience, it can be expected that greater familiarity or similarity will translate to a richer representation, which involves more associations formed through multiple interactions and in many different situations. This, in turn, will result in a more complex but also more accurate pattern of activity evoked by seeing the emotional state (for instance, distress) of the familiar individual.

In a later article, de Waal and Preston (2017) summarise further developments of the theory and emphasise that understanding of empathy should balance its emotional and cognitive aspects. They propose that empathy can be treated as an umbrella term for a set of distinct yet interrelated phenomena. These phenomena all rely on the PAM and are organised into three layers, from the most basic to the most complex. Due to the hierarchical organisation, in which higher layers originate from the lower, the framework

[3]"Not only do I see gravity and modesty and pride and courtesy and stateliness, but I feel or act them in my mind's muscles. This is, I suppose, a simple case of empathy, if we may coin that term as a rendering of *Einfühlung*" (Titchener, 1909)

[4]A detailed discussion of these themes is beyond the scope of this introduction, but briefly they were listed as follows: "There are functional differences between empathy and related concepts; empathy includes both cognitive and affective elements; the emotions of the target and observer are similar but not identical; other stimuli, such as imagination, can evoke empathy; a self/other distinction is maintained in empathy, although a degree of merging is necessary; empathy is affected by both trait and state influences; behavioural outcomes are not part of empathy itself; and finally, empathy is automatically elicited but is also subject to top-down controlled processes" (Cuff et al., 2016).

is referred to as the Russian doll model. The three layers of the model are summarised below.

The first layer includes motor mimicry (synchronisation of motor behaviours) and emotional contagion (emotional state matching between a target and an observer). One behaviour that can be classified under motor mimicry is yawn contagion, which has been observed in different species, from birds (e.g. budgerigars), through canines, to chimpanzees (de Waal & Preston, 2017). Although yawning can be seen as a mere physiological response, it can also convey internal or emotional state information and help build group synchrony. An interesting recent example can be found in lions (Casetta et al., 2021). Face and movement matching, which has been primarily studied in monkeys and apes (macaques, baboons, orangutans, and chimpanzees), are other examples of motor mimicry. Emotional contagion involves behaviours that are characterised by sharing the emotional state of another. For example, mice increase their pain responses when witnessing the pain of other mice, and prairie voles match the hormonal stress responses of their conspecifics (de Waal & Preston, 2017).

The second layer includes empathic concern (concern about another's state and attempts to ameliorate this state) and consolation (reassurance behaviour directed at a distressed party, such as a victim of aggression). Related behaviours have been predominantly observed in apes (chimpanzees, bonobos), which comfort others in distress by touching and hugging. However, consolation behaviours are also present in macaques, canines, elephants, voles, mice and rooks (de Waal & Preston, 2017). These behaviours are based on emotional contagion but require self-regulation (because simply immersing in the negative state of others is not sufficient to ameliorate it).

Finally, the third layer includes perspective taking (the capacity to take another's affective perspective: for example, understanding their specific situation and needs, separate from one's own, which still requires access to personal representations of the other's state) and targeted helping (assistance and care based on a cognitive appreciation of the other's specific need or circumstances). In humans, perspective taking is typically understood through performance in tests such as the false belief task. When a child is shown a candy box containing pennies and asked what another person would expect to find inside, children under about three years old would answer pennies when older children would correctly say candy (American Psychological Association, n.d.). Apes can pass similar tests (the subject looks in the direction where they expect another to search for a misplaced object: either to its current location or to the location last seen by another). However, in non-human animals, perspective-taking is more often inferred from helping behaviours. For example, when a dolphin lifts a companion to the surface to help them breathe, or a chimpanzee brings fruits to an elderly conspecific who cannot climb, they likely exhibit a combination of emotional contagion, self-regulation and perspective taking (de Waal, 2009; de Waal & Preston, 2017). However, when a rat liberates another from an uncomfortably tight compartment (Ben-Ami Bartal et al., 2011), it is not clear whether that involves perspective taking or merely trial-and-error learning motivated by emotional contagion.

It is worth noting that the interpretation of empirical findings may be ambiguous, as made evident by the last example. The findings may require careful examination be-

fore they are classified in the empathy framework (Adriaense et al., 2020). For instance, one experiment (Yamamoto et al., 2012) investigated pairs of chimpanzees separated by a barrier: one was given a box with several tools, while the other needed one of these tools (stick or straw) to obtain a juice reward. When chimpanzees were able to see through the barrier, in most cases, they offered the correct tool in their first attempt to help. However, they only did so reactively, after visible requests for assistance, rather than proactively. This behaviour can be interpreted as proof of a cognitive ability to understand the other's goals (perspective taking) and a display of targeted helping, although without the emotional basis (emotional contagion) presupposed by the Russian doll model. This observation led Yamamoto (2017) to propose an alternative model of empathy. This model also treats empathy as an umbrella term and divides it into three components: matching with others, understanding of others, and prosociality. Crucially, the components are considered to be independent of each other. Although they may overlap, they are not assumed to rely hierarchically on each other. In this framework, targeted helping originates from a combination of prosociality and understanding of others, while emotional contagion belongs solely to the matching with others category.

Going back to the Russian doll model, its innermost layer seems most relevant for the current work. In one of the previous sections, I suggested that vicarious fear conditioning can be treated as a learning process. At the same time, it is an empathic process: watching the demonstrator experiencing aversive stimulation involves perceiving their emotional expressions, which should lead to matching their emotional state. In this way, the emotional contagion mechanism is involved in fear conditioning through observation. Therefore it is reasonable to expect that a higher propensity for emotional contagion (or empathy in general) may result in more effective fear conditioning. Moreover, since the PAM suggests that familiarity and similarity modulate empathy (through greater richness and accuracy of representations), it is likely that these two factors should also contribute to observational fear conditioning.

Two human studies attempted to examine the relationship between empathy and observational fear conditioning directly. In the first of these studies (Olsson et al., 2016), three groups of participants received instructions to either enhance their empathic responses, diminish them, or received no instruction at all. The group instructed to enhance their empathy showed the strongest vicarious fear learning measured by skin conductance responses. Moreover, the responses correlated with trait (i.e. questionnaire-measured) empathy; however, only in the group instructed to empathise actively. Per contra, another study failed to replicate the relation between fear conditioning and trait empathy, using a single group with high-empathy instruction (Williams & Conway, 2020). Although conditioned responses were overall present, they did not correlate with any of several empathy-related questionnaires. Another two studies, also based on skin conductance responses, aimed to investigate the influence of social and racial similarity. One study (Golkar et al., 2015) found stronger fear conditioning when observing a person from the same racial group. The other experiment (Golkar & Olsson, 2017) found an additional influence of a social group (specifically, sports club supporter group). Within the same racial group, supporters of the same team exhibited the highest conditioned responses, and supporters of rival teams exhibited the lowest conditioned

responses.

1.6 Towards ecological validity of observational fear conditioning

For humans, social learning happens through spontaneous interactions in their natural environment (e.g. home, classroom, city streets or summer campsite). Contrary to that, 'traditional' research is carried out in laboratories, using simplified models and precisely controlled stimuli. Therefore, finding an appropriate experimental paradigm for social neuroscience can be challenging. An ideal experiment should strike a balance between ecological validity (i.e. being similar to real-world experiences) and tight laboratory control (Bottenhorn et al., 2019; Matusz et al., 2019). Speaking broadly, recognition of a need for ecologically valid social experiments is not new. Already forty years ago, Neisser criticised the tendency to focus on passive observation or judgement protocols, concluding that "When people are genuinely engaged with each other, nobody stops to give grades" (Neisser, 1980). However, recent technological developments have enabled experimental creativity to a large degree, be it the existence of network protocols for synchronising different streams of data (such as LabStreamingLayer, github.com/sccn/labstreaminglayer) or simply the availability of inexpensive, high-quality digital cameras for video capture or streaming.

Several recent studies on vicarious fear conditioning in humans utilised the observational fear conditioning protocol (Haaker, Golkar et al., 2017). This protocol consists of two stages: observational learning and direct-expression, which correspond to the acquisition and test phases of classical conditioning. First, the participant (observer) watches a video recording of another person (demonstrator) being presented with two neutral CS; one is paired with aversive electric stimulation, and the other is not. In this situation, the demonstrator's reactions to the stimulation serve as the observational US for the observers. Next, the participant's expression of learned threat is measured as the differential conditioned response (e.g. skin conductance) to the direct presentation of the same CS as were presented to the demonstrator (in the absence of the demonstrator). Aversive stimulation is never applied to the observer throughout the entire procedure to ensure that resulting threat acquisition is due to observation only.

One way to perform an experiment designed in such a way is to use video recordings of a demonstrator as social stimuli. Indeed, this is how the observational fear conditioning protocol has been formulated (Haaker, Golkar et al., 2017). On the one hand, this approach has clear advantages: it offers a high degree of control over experimental stimuli and allows using a single demonstrator across multiple observers. On the other hand, viewing videos limits the naturalness of the social situation. Therefore, conducting a similar experiment in real-time with pairs of participants may offer alternative benefits, such as increased sense of involvement, higher ecological validity and the possibility to design the experiment based on an existing relationship between the demonstrator and the observer.

In practice, fear conditioning studies involving participant dyads are scarce. To my knowledge, apart from the experiments presented in this thesis, only two dyadic adaptations of the observational fear conditioning protocol described above have been published[5]. Both were carried out independently and in parallel to the current work and consequently were methodologically different. One study (Pärnamets et al., 2020) investigated pairs of participants, unfamiliar with each other, sitting in the same room and alternating between the roles of demonstrator and observer during the experiment. Its main focus was on the physiological synchrony of electrodermal signals and its ability to predict threat learning. In another study (Pan et al., 2020), with further-reaching modifications, the observational aspect was introduced to investigate fear extinction rather than acquisition. First, both participants individually completed an acquisition stage while separated by a movable wall, which was then removed for the extinction stage (during which mirrors were provided to allow them to see each other's reactions without taking their eyes off computer screens). Both studies used pairs of participants who were not familiar with each other before the experiment, and both used electrodermal activity measurements as the only measure of conditioning outcome.

1.7 Measures of fear conditioning in humans

Experiments described in the thesis used several methods to quantify the effects of observational fear conditioning, relying both on functional magnetic resonance imaging (fMRI) as well as on psychophysiological measures: electrodermal activity (EDA) and fear-potentiated startle (FPS). The rationale for their inclusion and their essential properties are summarised below.

Self-reports

Perhaps the most basic way of measuring fear conditioning relies on participant's declaration, in the form of either a free report, a questionnaire or a scale (Ojala & Bach, 2020). One approach is to collect expectancy ratings: during the experiment, participants can be asked about their expectation to receive (or not) an aversive stimulus at a given moment (Boddez et al., 2013). This method is usually applied in studies investigating learning trajectories over time. Another approach may utilise valence ratings of the conditioned stimuli, given at different moments during the experiment, in order to evaluate the change in their perceived valence due to conditioning, (e.g. Blechert et al., 2008). Yet another method is to test contingency knowledge by asking participants after the experiment whether one of the conditioned stimuli predicted the unconditioned stimulus, (e.g. Weidemann et al., 2016). When testing contingency knowledge, a hierarchical procedure rather than a single forced-choice question should be used to avoid chance responses. The post-experimental contingency knowledge assessment was

[5] In this context, one of the seminal studies for observational fear conditioning (Bandura & Rosenthal, 1966), can be seen as covering a middle ground: the observation was indeed carried out live, but an experimenter's confederate, rather than an external participant, served as a demonstrator.

selected to be used in both experiments described in this thesis to assess the degree of declarative learning.

Electrodermal activity (EDA)

Measuring electrodermal activity, i.e. changes in the skin's electrical conductance is arguably the most classical method used in human fear conditioning studies. It has been employed since the 1930s and has been suggested as a primary measure in the observational fear conditioning protocol (Haaker, Golkar et al., 2017). Although sweating is primarily a thermoregulatory mechanism, it can also provide insights into cognitive states.

From a physiological standpoint, eccrine sweat glands are innervated by sympathetic sudomotor nerve fibres. The phasic firing of these fibres causes opening of the sweat glands and secretion of sweat, which results in an increase of skin conductance, referred to as the Skin Conductance Response (SCR). Therefore, in a most general sense, SCR can be considered a marker of sympathetic arousal. Such a reaction is not specific to threat and can be elicited by stimuli of different valence. However, in fear conditioning, higher SCRs are elicited by CS+ compared to CS–. Moreover, cognitive-computational studies investigating trial-by-trial changes of the CR suggest that SCRs most likely reflect a mixture of US prediction and its uncertainty (Ojala & Bach, 2020).

The most common method of quantifying SCRs relies on peak scoring, i. e. calculating the difference between a peak and a preceding foot point observed in the time course. A typical reaction has an onset between 1 and 4 seconds after a stimulus and a peak between 0.5 and 5 seconds later (Boucsein et al., 2012). However, exact criteria for peak scoring have been often debated, (e.g. Pineles et al., 2009). Alternative, model-based approaches have also been proposed. One such approach relies on the fact that the SCR is highly prototypical, to such an extent that a canonical response function can be established empirically and used in a general linear convolution framework (Bach et al., 2009, 2010). Another approach uses a generative model of SCR, based on intraneural recordings (Gerster et al., 2018), to estimate the most likely parameters (onset, latency and dispersion) of the sudomotor nerve bursting given the observed skin conductance data. Apart from the phasic SCR, slower tonic changes can provide additional information about the general level of arousal and can be quantified in terms of skin conductance level (SCL) or frequency of spontaneous fluctuations.

Since skin conductance measurement is relatively simple, it can be combined with other methods. Consequently, the psychophysiological experiment combined SCR with FPS. In the neuroimaging experiment, skin conductance data was recorded during fMRI acquisition, thanks to an MR-compatible recording system.

Fear Potentiated Startle (FPS)

Startle reflex is a fast defensive reaction to sudden, intense stimuli and includes several components, such as head withdrawal, shoulder elevation and closing of eyes. In human experiments, the eyeblink component is recorded from the *orbicularis oculi* muscle

through surface electromyography (EMG). The startle reflex can be elicited by a presentation of a short, loud sound, typically a white noise burst, and in the literature is referred to as startle eyeblink or acoustic startle response. Notably, although the response is seemingly very primitive, its strength can be modulated by external factors.

The primary neural pathway controlling the acoustic startle response leads from the cochlear root nucleus through the caudal pontine reticular nucleus to motor neurons. Crucially, the caudal pontine reticular nucleus receives, among others, excitatory input from the central amygdala, which is a primary source of response enhancement due to amygdalar activation (Koch, 1999; Kuhn et al., 2020). In fear conditioning, the startle response is higher during CS+ compared to CS–. Consequently, the term fear-potentiated startle (FPS) is used. The startle response can be probed during the presentation of task-relevant stimuli, as well as during intertrial intervals, which provides an additional baseline condition (Lonsdorf et al., 2017; Ojala & Bach, 2020). The startle response is considered a reliable marker of amygdala-dependent learning, which additional factors, including declarative memory, can nevertheless influence; see (Ojala & Bach, 2020) for a detailed discussion. Fear potentiated startle was used in the psychophysiological experiment described in this thesis.

Functional Magnetic Resonance Imaging (fMRI)

Magnetic Resonance Imaging was the primary method of investigation used in the current thesis. It is a non-invasive method of imaging brain structure (structural MRI) and function (functional MRI, fMRI). The fMRI data are obtained as a series of images, with a spatial and temporal resolution in the order of millimetres and seconds, respectively.

Image intensity is produced by magnetic field inhomogeneities caused by the presence of deoxyhemoglobin in red blood cells. Consequently, fMRI directly reflects blood oxygenation, as well as cerebral blood flow and volume (Logothetis & Pfeuffer, 2004). Intensity values obtained for any given voxel (volume unit) in the image create a time series called the Blood Oxygenation Level Dependent (BOLD) signal. The BOLD signal, per se, is an indirect measure of neural activity; however, it is tightly coupled to underlying neural events. Animal experiments combining MRI with invasive electrophysiology have shown that local field potentials (the low-frequency component of extracellular field potential, reflecting synchronous, primarily postsynaptic, activity in neural populations) predict the BOLD response, and, consequently, that the BOLD signal reflects synaptic input and processing in a given area (Logothetis et al., 2001; Logothetis & Pfeuffer, 2004). The primary effect observed in the BOLD signal is called the haemodynamic response: an increase of the signal intensity, peaking approximately 4 - 6 seconds following stimulus presentation, followed by a slower decay and an undershoot, and returning to baseline approximately 12 - 20 seconds after the onset (Poldrack et al., 2011). This response is caused by an increased presence of oxygenated blood, which overcompensates the initial energy demand of active neurons.

fMRI analysis - activation framework

Arguably the most common approach to the analysis of fMRI data focuses on identifying activations, that is, changes in the BOLD signal caused by some experimental manipulation. The activations are typically analysed by contrasting different conditions: a voxel (or an area) is considered to be activated if it exhibits a higher response in one condition compared to another and if the difference is statistically significant.

In most experimental paradigms, the haemodynamic response can be treated as linear and time-invariant: it scales linearly with the intensity of the underlying stimulus, responses to separate stimuli are additive, and response to a stimulus shifted in time is likewise shifted. Given that a typical experiment involves several conditions, the overwhelmingly dominant approach to fMRI data analysis is based on the general linear model (GLM), which explains the BOLD signal in each voxel as a linear combination of responses to different stimuli, confounding factors and normally distributed noise (Poldrack et al., 2011; Poline & Brett, 2012). On the implementation side, the predicted time series is first constructed for each condition by convolving the time course of stimulation with a function representing the canonical haemodynamic response. Next, all predicted time courses are entered into a linear regression model (together with additional regressors, such as the subject's head motion, which may confound the signal), and the model is fitted to the observed data. Therefore, a response to a given condition is summarised by a model parameter (which essentially describes how well the observed time course matches a predicted response), and contrasts can be defined based on these parameters.

One important aspect of analysing the fMRI data is related to the fact that the above procedure is performed for each voxel separately (which is referred to as a mass-univariate approach) and necessitates a statistical adjustment for multiple comparisons. While an image may contain about 200 000 voxels, they are not entirely independent, as there is a significant spatial correlation. Therefore, the statistical analysis aims to control for a familywise error rate (FWER or FWE; that is, for the chance of finding false positives anywhere in the image). The threshold for a significant finding may be inferred from the smoothness of the data or, less commonly, obtained through permutation testing. It can be calculated either in terms of either individual voxel values (voxel-based FWE) or the number of contiguous voxels exceeding a certain threshold (cluster-based FWE) which are unlikely to appear in random data of a given smoothness (Brett, 2016; Nichols & Hayasaka, 2003; Poldrack et al., 2011).

fMRI analysis - connectivity framework

Another conceptually different approach to analysing fMRI data is identifying functional connectivity, that is, the existence of a relationship between BOLD activity observed in different regions. A fundamental notion for functional connectivity is that a temporal correlation between physically distant brain regions can indicate that they are functionally related. A seminal study (Biswal et al., 1995) demonstrated the existence of such a relationship between the left and right sensorimotor cortex in resting-state

fMRI (i.e. in the absence of any task). Since then, numerous studies of resting-state functional connectivity have shown that the brain can be segregated into a limited set of networks, which are functionally relevant and highly consistent across individuals (Biswal, 2012; Fox & Raichle, 2007; Power et al., 2011; Smith et al., 2009). However, individual differences from the general organisation can also become apparent, especially with extensive acquisition (multiple scanning sessions, resulting in several hours of data) from individual subjects (Gordon et al., 2017).

The psycho-physiological interactions (PPI) framework is one of several kinds of connectivity analysis methods. It is specifically dedicated to investigating changes in connectivity during a task (i.e. task-related connectivity). In an experiment with A and B conditions, a PPI effect can be observed in a region that exhibits similar activity to a selected "seed" region of interest in condition A but not in condition B. The target region can be selected a priori (to examine how strong the effect is for a pair of regions), or a seed-to-voxel analysis can be conducted to find all regions that exhibit a significant effect. Formally, the expected signal is modelled as a product of a psychological variable (which is a categorical indicator, equal to 1 for the presence of condition A, -1 for the presence of condition B, and zero otherwise) and the physiological activity (i. e. BOLD signal[6]) of the selected seed. The resulting interaction term is entered into a GLM analysis, analogous to that of task activations. Additionally, psychological and physiological variables are included as regressors of no interest to guarantee that detected effects are not due to simple coactivation or correlation regardless of task (Poline & Brett, 2012).

This formulation means that the expected time course is characterised by a similarity to the seed in condition A and dissimilarity in condition B[7]. As a result, detecting an effect in a given brain region can be interpreted as a task-dependent change in connectivity with the seed region - in other words, an interaction of the psychological and physiological factors, going above and beyond their main effects (Di et al., 2020; O'Reilly et al., 2012).

The PPI analysis was used in the neuroimaging experiment to elucidate further the interactions between the brain structures involved in observational fear conditioning, primarily in the context of the presence or absence of the vicarious unconditioned stimulus.

1.8 Neural correlates of fear conditioning in humans

Fear conditioning has been extensively studied in humans using fMRI. A recent meta-analysis (Fullana et al., 2016), comprising 27 independent studies published between 1998 and 2013, found consistent evidence of brain activation in a set of brain regions, which can be considered an extended 'fear network'. Specifically, meta-analytic activations for

[6]The haemodynamic signal is temporarily delayed compared to the underlying neural activity. In some implementations, notably that of SPM software, the BOLD signal is first deconvoluted with the HRF, and the interaction term is calculated in terms of probable underlying neural activity, rather than the BOLD signal itself, which is more suitable for short events.

[7]In the ideal time course. During analysis, a time course defined in this way should also match signals with greater (correlational) similarity in condition A and lesser in condition B.

the (CS+ > CS–) contrast were found in the following locations: anterior insula (AI) extending to frontal operculum, ventral striatum and thalamus, dorsal anterior cingulate cortex (dorsal ACC, dACC), pre-supplementary and supplementary motor cortex, precuneus (both dorsal-anterior and ventral-posterior parts), parietal operculum, dorsolateral prefrontal cortex, lateral premotor cortex, lateral cerebellum, and a set of smaller subcortical regions. Particularly interesting is the involvement of the anterior insula and the dorsal ACC, regions with extensive anatomical connectivity, which are also linked to the production of subjective feelings and coordination of responses to internal and external events (Medford & Critchley, 2010). Notably, the meta-analysis did not find robust amygdala involvement; however, this might be due to the transient nature of its responses (primarily to the US) or the specifics of the fMRI fear conditioning protocols (Fullana et al., 2016). Nevertheless, the amygdala has been considered to be an essential part of the fear network. This role is supported by human fMRI studies and animal literature (LeDoux et al., 1990; Phelps & LeDoux, 2005). The amygdala is considered a place for the convergence of CS and US sensory inputs forming the CS–US associations.

Several fMRI studies in which the observational fear conditioning protocol has been applied extend these findings to fear conditioning through observation. Two of these are particularly informative for the current work. The first confirmed the involvement of the amygdala, AI, and ACC in both acquisition and expression of socially learned fear. The second study (Lindstrom) used a within-subject design to compare observational and direct (self-experienced) fear acquisition and again demonstrated the involvement of these three structures in both conditions. Moreover, the amygdala was identified as the most probable input for the US in the direct condition and the AI in the social condition. Additionally, the AI was more connected to the temporoparietal junction (TPJ) during the social condition. These studies suggest that the amygdala, AI, and ACC can be considered a core fear network.

Other experiments provided further evidence. Haaker, Yi et al. (2017) used a pharmacological manipulation to demonstrate that the opioidergic system, also engaged in direct threat learning, is relevant for observational fear conditioning. This study also demonstrated activation of the amygdala and periaqueductal grey and a functional connection between periaqueductal grey and superior temporal sulcus (STS). Phelps et al. (2001) demonstrated that a verbally instructed threat can also engage the amygdala and AI. Golkar et al. (2016) described the role of the ventromedial prefrontal cortex in vicarious extinction learning. More recently, Silvers et al. (2021) showed higher responses in the left amygdala in children learning from parents compared to unfamiliar adults.

A recent review of social value-based learning (Olsson et al., 2020) indicates that social learning of threats mainly shares its neural mechanisms with self-experienced learning. Additionally, social threat learning recruits regions specific for social cognition, such as the dorsomedial prefrontal cortex, temporal pole, TPJ, and STS (Olsson et al., 2020). Notably, the latter two structures have been described as primary elements of a social perception model (Yang et al., 2015) in which the TPJ is responsible for higher-level processes related to the theory of mind (perspective taking and evaluation of others' mental states), while the posterior STS integrates lower-level information related to social perception and action observation.

Finally, a comment on terminology is required. Several conventions exist for dividing the anterior cingulate cortex into subregions, and different nomenclatures are used in the literature (van Heukelum et al., 2020). One convention divides the ACC into ventral (vACC) and dorsal (dACC) parts. The vACC is located near the genus of the corpus callosum, while the dACC is located dorsal to the body of the corpus callosum. Another conversion uses the name ACC to describe only the vACC and considers dACC as part of the mid-cingulate cortex (MCC) instead. Therefore, anterior MCC (aMCC) and dACC refer largely to the same location (van Heukelum et al., 2020). In the thesis, the term aMCC will be used preferably to dACC, following the region of interest definitions (Alcalá-López et al., 2018) selected for part of the analyses, but the term ACC will be understood broadly.

1.9 Research goals and questions

As discussed above, previous studies provided information about the similarities between observational and self-experienced fear conditioning and clues about potential factors modulating the process. In particular, observational fear conditioning is considered tied to empathy, but the extent to which empathy influences fear conditioning is not clear. Furthermore, the effects of familiarity between the demonstrator and the observer have not been investigated. It can be hypothesised that familiarity should enhance observational fear conditioning due to higher empathy towards familiar demonstrators. However, it is unknown whether the influence of empathy and familiarity would indeed be reflected when examining pairs of friends, especially at the level of brain activation. Choosing and developing an experimental protocol to investigate these aspects was particularly important. The design choices aimed to enhance the ecological validity and social aspects of existing protocols. Ultimately, the procedure applied here should be applicable to study various factors other than friendship. Based on these considerations, the experiments described in this thesis were designed to achieve the following goals:

1. Create an expanded version of the observational fear conditioning protocol (Haaker, Golkar et al., 2017), in which pairs of participants are involved in the experimental procedure together, which could be applied in psychophysiological and neuroimaging experiments.

2. Test whether fear conditioning in the modified protocol is effective in pairs of friends in a proof-of-concept experiment. Successful fear conditioning was expected to be reflected in two physiological measures: electrodermal activity and fear-potentiated startle.

3. Test, in a separate experiment, whether friendship enhances observational fear conditioning by comparing physiological responses and brain activation patterns directly between participants who observed their friends and participants who observed persons they did not know. Evaluate the main effects of the task (by

analysing both groups of observers together) and perform between-group comparisons, aiming to:

a) investigate the task-related activation during both observation and direct-expression stages

b) investigate the task-dependent connectivity between regions relevant to both fear and social perception

The description of the first (psychophysiological) experiment has been previously published in (Szczepanik et al., 2020). The second (neuroimaging) experiment has appeared in a preprint (Kaźmierowska et al., 2021). Here, the descriptions of both experiments are expanded and they are discussed in the broader context of the research project.

Chapter 2

Methods

This chapter introduces the methods used in two consecutive experiments upon which the thesis is based. The first experiment was primarily psychophysiological and was carried out at the Faculty of Psychology, University of Warsaw. The second (neuroimaging) experiment used fMRI as its core method. It was conducted at the Laboratory of Brain Imaging at the Nencki Institute of Experimental Biology of the Polish Academy of Sciences. Protocols used in the studies were approved by the Ethics Committee of the Faculty of Psychology at the University of Warsaw (decision from 28 November 2017) in accordance with the Code of Ethics and Professional Conduct of the Polish Psychological Association (Polskie Towarzystwo Psychologiczne) and the Ethics Code of the American Psychological Association (APA).

2.1 Shared methods

Behavioural measures

McGill Friendship Questionnaire

The McGill Friendship Questionnaire - Respondent's Affection[1] (Mendelson & Aboud, 1999) was used for initial screening during the recruitment of volunteers. It contains 16 items describing feelings for a friend and satisfaction with the friendship. Positive statements, such as "I want to stay friends with ... for a long time", are responded to along a 9-point scale ranging from -4 (very much disagree) to 4 (very much agree). Consequently, the total score can be between -64 and +64 points. Responses were collected online, and one item (no. 9) was omitted due to human error.

State and Trait Anxiety Inventory

The State and Trait Anxiety Inventory (STAI; Spielberger et al., 1983; Spielberger et al., 2012) was used to control the participants' state and trait empathy during the experiment. It consists of two 20-item subscales, STAI-State and STAI-Trait, rated using a

[1] Polish translation was prepared for the current studies by A. Kaźmierowska, P. Pączek & A. Schudy.

4-point scale. Higher scores indicate a greater level of anxiety. The STAI-State was completed twice by each participant (before and after the experiment), and STAI-Trait was completed once (at the end).

Basic Empathy Scale for Adults

The Basic Empathy Scale for Adults[2] (BES-A; Carré et al., 2013; Jolliffe & Farrington, 2006) questionnaire was used in the psychophysiological experiment to control trait empathy. It contains 20 items (such as "I usually feel calm when other people are scared" or "It is hard for me to understand when my friends are sad"), rated by the participants using a 5-point scale from "strongly disagree" to "strongly agree". It can be decomposed into either two-factor (affective and cognitive empathy) or three-factor structure (emotional contagion, cognitive empathy, emotional disconnection). It was completed by the participants at the end of the experiment.

Empathic Sensitiveness Scale (Skala Wrażliwości Empatycznej)

The Empathic Sensitiveness scale (original title: Skala Wrażliwości Empatycznej, SWE; Kaźmierczak et al., 2007) was used to measure trait empathy in the neuroimaging experiment. It was selected to replace BES-A for two reasons: it was validated in Polish, and its items were deemed more accurate. The questionnaire is based on the Interpersonal Reactivity Index (Davis, 1983), but far-reaching changes have been made during the development of SWE as its Polish adaptation. It contains 28 statements, which reflect three components of empathy: two emotional (empathic concern, personal distress) and one cognitive (perspective taking). The answers are given on a 5-point scale. Participants completed this questionnaire at the end of the experiment.

Assessment of stimulus contingency

A questionnaire adapted from (Weidemann et al., 2016) was used to assess CS-US contingency knowledge. In this questionnaire, participants were asked about the observation stage. It started with a question of whether the participant could predict when the shocks would occur, conditionally followed by an open-ended question ("if so, how"). Next, participants gave percentage ratings of shock occurrence for each stimulus (CS+, CS–, fixation symbol). Finally, participants were asked to select only one stimulus paired with the shock (CS+, CS or fixation symbol). The questions were presented sequentially on a computer. Although the questionnaire referred to the observation stage, it was completed by the observer after the experiment.

Evaluation of the demonstrator and the observational US

This questionnaire, published as supplementary material to the observational fear conditioning protocol (Haaker, Golkar et al., 2017), was used to quantify the observation

[2]Polish translation was prepared for the current studies by A. Kaźmierowska & A. Schudy.

experience. It contains five items in which the participant is asked to rate the perceived demonstrator's discomfort, strength and naturalness of reactions, own empathy, and identification with the demonstrator. Items are rated from 0 (not at all; very poor) to 9 (very much; very strong). Observers filled this questionnaire at the end of the experiment.

Stimuli

Two coloured squares, blue and yellow, covering approximately half of the screen height, were used as conditioned stimuli. The assignment of colour to CS+/CS– was counterbalanced across participants. Each CS was presented for 9 seconds. A centrally located '+' symbol served as a fixation point and was used during intertrial intervals (ITI). The ITI lasted between 10 and 15 seconds, randomised between trials. All symbols were displayed on either black (psychophysiological experiment) or grey (neuroimaging experiment) background.

Cutaneous electrical stimulation applied to the upper ventral part of the right forearm, consisting of five 1 ms unipolar pulses with 200 ms interval, was used as the US. This location, above the flexor carpi radialis muscle, was chosen for stimulation because it could produce flexion at the wrist in most subjects, visible to the observer. Stimulation intensity was individually adjusted to be unpleasant but not painful (see demonstrator's preparations below). In reinforced trials, the stimulation started 7.5 seconds after the CS onset. Consequently, a reaction to this stimulation co-terminated with the CS.

Observational fear conditioning protocol

The experimental procedure was based on the observational fear conditioning protocol (Haaker, Golkar et al., 2017). The original protocol used video recordings, prepared in advance and displaying an actor, as stimuli for the observer. Moreover, it was formulated primarily for EDA measurements. Several significant modifications were made for the experiments described in the thesis. First, pairs of friends were recruited to participate in the study in real-time, and the observation was conducted through video streaming. One of the friends was the demonstrator, and the other was the observer. Second, additional measures of fear conditioning were included: the psychophysiological experiment used EDA and FPS, and the neuroimaging experiment used fMRI and EDA. Consequently, the number and duration of stimulus presentations were extended to accommodate these methods. Third, in the neuroimaging experiment, the video stream was additionally recorded to allow between-group comparisons, where a separate group of unfamiliar observers watched the same demonstrators. The procedure consisted of two stages, described in detail below.

Observational learning stage

In the observational learning stage, the demonstrator performed a differential conditioning task while the observer watched. Before the task, the observer was informed

that the person they will watch will be presented with visual stimuli and may experience electrical stimulation. The observer was also instructed that they would perform the same task in the next part of the experiment. Crucially, these instructions contained no information about the potential relationship between the stimuli.

After the instructions, video streaming was turned on, and stimulus presentation followed. 48 trials were presented: 24 CS+ (of which 12 were reinforced) and 24 CS–. Only the demonstrator experienced electrical stimulation. The trial order was pseudorandom with a limitation that the same CS can occur no more than twice in a row. The first and the last CS+ had to be reinforced. The camera was positioned to capture the demonstrator's face, upper body (including the stimulated hand), and computer screen, and the video was streamed without sound.

Direct-expression stage

After observational learning ended, the video stream was turned off, and the observer was informed that they were about to take part in the same task as they had watched. After the instruction, the direct-expression stage started. On their computer screen, the observer viewed the same conditioned stimuli as used in the previous stage. Crucially, the electrical stimulation was not applied to ensure that the observer has only observational experience with this stimulus. The direct-expression stage was shorter than the previous one and consisted of 24 trials, 12 CS+ and 12 CS–. The trial order was pseudorandom, and given CS could not repeat more than twice.

Procedure

Before the experiment

Upon arrival at the laboratory, participants were briefly introduced to the details of the experimental procedure, provided informed consent and filled safety questionnaires. The questionnaires were related to electrical stimulation and, in the neuroimaging experiment, MR safety. The roles of observer and demonstrator were then assigned by participants choosing from two colour-coded envelopes. In this way, the assignment was immediately known to the experimenter (envelope colour) but not the participant (a note inside the envelope: either "demonstrator" or "observer"). Consequently, participants could remain naive to the specifics of their task at the beginning of the experiment when baseline measures (e.g. questionnaires) were collected.

Next, the participants were separated and placed in two adjacent rooms. At this point, they filled initial questionnaires. From this point onwards, they proceeded through the experiment separately.

Demonstrator's preparation and stimulus intensity adjustment

First, the demonstrator had stimulation electrodes and sham recording electrodes attached. Although physiological recording from the demonstrator was not possible due

to technical limitations, they wore the same electrodes as the observer. Using sham recording electrodes helped create an impression that both participants were about to perform identical tasks. Subsequently, the demonstrator was asked to open the previously drawn envelope and was informed about the upcoming task. Unlike the observer, they were given complete information about the procedure, including which stimulus would be paired with an electric shock. They were asked to react to the shocks in a manner that would be natural but also unambiguous to the observing person (such as a facial grimace) and to remain calm when no shocks would be applied. An example video of a person reacting to a shock was shown. These instructions were provided to harmonise the behaviour of different demonstrators. Instructions were included as a consequence of previous pilot experiments, in which several observers reported that they did not notice the stimulation onset.

Following the instructions, the stimulation intensity was adjusted. The intensity was increased stepwise. The participant was asked to rate each stimulus using a scale with eight items, ranging from imperceptible (1) to painful (8). The item "very unpleasant, but not painful" (6) was the target. Finally, the demonstrator was slightly angled relative to the computer screen, and a camera was placed to their side. In this arrangement, the camera could capture all relevant cues: face, stimulated hand, and computer screen. The observer did not see these preparations.

Observer's preparation and tasks

Meanwhile, the observer had recording electrodes and sham stimulation electrodes attached. Next, they opened the previously drawn envelope and were informed about the upcoming tasks. Their instructions were vague about stimulus pairing, unlike those given to the demonstrators. The observers did not perform stimulus intensity adjustment. The preparation was followed by observational learning and direct-expression described above.

After the experiment

After completing their last part of the procedure, each participant filled in post-experimental questionnaires (see behavioural measures). This was after the observational learning for the demonstrator, and for the observer, this was after the direct-expression stage. Finally, both participants were debriefed about the study.

2.2 Psychophysiological experiment

Participants

Thirty-five pairs of male friends aged between 18 and 27 (M=21.4, SD=2.2) participated in the study. Two inclusion criteria were used to define pairs of friends: knowing each other for at least three years and obtaining at least 50% of the maximum score (by each participant) in the McGill Friendship Questionnaire (Mendelson & Aboud, 1999). Ad-

ditionally, only participants declaring themselves heterosexual (to restrict the relationships to non-romantic male friendship and reduce sample variability) were recruited. Exclusion criteria included colour blindness, the presence of neurological disorders or conditions precluding electrical stimulation, psychoactive medication usage, and being a student of psychology or cognitive science (to exclude participants familiar with fear conditioning procedures). Each participant received financial remuneration of 60 PLN (~15 EUR) for their participation.

Stimuli

In addition to the visual CS and aversive electrical stimulation, bursts of white noise were used to elicit the startle reflex. A startle probe had a 50 ms duration and 80 dB(A) intensity, which raised near-instantaneously. The loudness was chosen to remain in line with recommendations (Blumenthal et al., 2005) while not being overly uncomfortable. The goal was to limit interference with fear conditioning (essential since the observers who heard the startle never experienced electrical stimulation directly).

A time window for probe onset was between 6–7 seconds relative to CS onset (if presented during CS) or between 2–4.5 seconds relative to the start of ITI (if presented during ITI). The time windows were chosen based on two considerations. First, presenting startle probes late during CS and early during ITIs meant that the noise burst would not confound a skin conductance response to CS onset. Second, startle probes were presented close to the potential US, as such presentation is the most sensitive in distinguishing threat and safety cues (Weike et al., 2008). Moreover, the timing during CS was similar (within 0.25 s) to that used in another study combining skin conductance and startle responses (Hamm & Vaitl, 1996).

Procedure and tasks

Because this experiment included startle reflex measurement, its procedure had some unique elements. Startle probes were delivered only to the observer (because only their responses were being recorded). However, both participants wore headphones to make them look similar. The observers were instructed to ignore the probe sound and informed that it was irrelevant to the task. Two short stages, startle habituation and resting-state, were also added.

During startle habituation, ten startle probes were presented in random 15–25 second intervals. Such habituation helped reduce startle probes' aversiveness and consequently prevented initial reactions during the following task from being exaggerated. The habituation was performed before the resting state and the observational learning stage (task instructions separated the two parts). During the resting-state signal acquisition, used to assess the baseline skin conductance level, a fixation point was displayed on the screen, and 12 startle probes were presented in random 10–20 second intervals, resulting in a total average duration of 6 min 20 s.

During the observational learning and direct-expression stages, startle probes were delivered during half of CS presentations and a quarter of ITIs, resulting in an equal

number of probes per condition (CS+ / CS– / ITI). The number and duration of trials and their timing were the same as described in shared methods. A schematic representation of the experimental design is shown in Figure 2.1. Stimulus presentation was controlled using Presentation v 19.0 (Neurobehavioral Systems, Berkeley, CA, USA).

A security camera was used to transmit a high definition video of the demonstrator, without sound, over a Serial Digital Interface (SDI) connection. The camera was connected through an SDI to HDMI converter directly to the observer's screen, and inputs were toggled using the screen button panel, making the setup simple and robust. The SDI signal is sent through concentric cables for tens of metres without loss of quality, making this solution adaptable to different laboratories. However, recording from the camera for further evaluation or use is not possible without additional components.

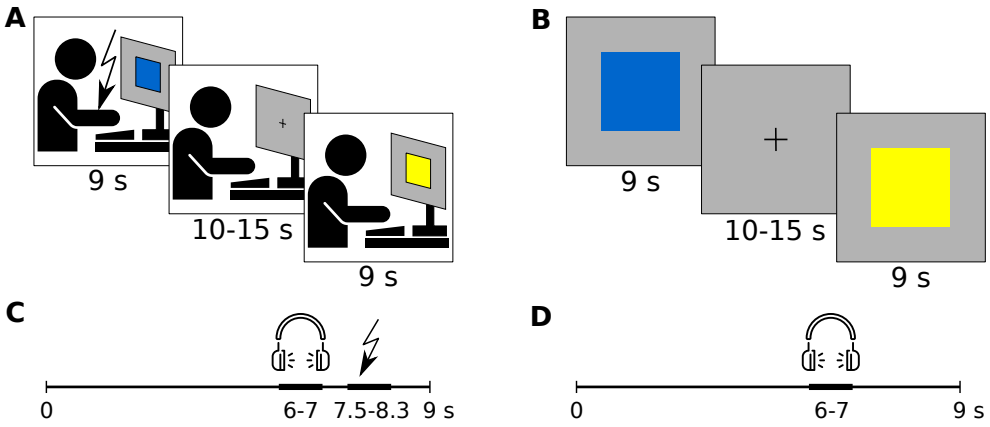


Figure 2.1: Trial sequence and within-trial timings for observational learning and direct-expression stages (as seen by the observers). In this example, the blue square is the CS+, and the yellow square is the CS–. (A) Trial sequence in the observational learning stage (half of CS+ were reinforced). (B) Trial sequence in the direct-expression stage. (C) Timing windows within CS: the 50 ms startle probe could occur between 6 and 7 s, the electrical stimulation could be applied to the demonstrator in a series of five pulses, equally spaced from 7.5 to 8.3 s. (D) The timing window for the 50 ms startle probe presentation during CS in direct-expression, electrical stimulation was not applied.

Physiological recording and stimulation

Physiological signals were recorded only from the observer, using the Biopac MP160 system equipped with EDA100C and EMG100C amplifier modules and AcqKnowledge software (Biopac Systems Inc., Goleta, California, USA). The sampling rate was set to 2 kHz.

For EDA recording, a 10 Hz low-pass hardware filter was enabled. Two Ag-AgCl electrodes with 6 mm recording diameter (TSD203) filled with electrode paste (GEL 101) were placed on distal phalanges of the right index and middle fingers using velcro straps.

For EMG recording (of the startle response), 1 Hz high-pass and 500 Hz low-pass hardware filters were used. Two Ag-AgCl electrodes with 4 mm recording diameter (EL254) filled with electrode gel (Signa Gel, Parker Laboratories Inc.) were placed under the participant's right eye, over the orbicularis oculi muscle. One of the electrodes was positioned in line with the pupil in forward gaze, the other 1–2 cm laterally. An additional ground electrode was placed on the upper central part of the participant's forehead, just below the hairline.

Cutaneous electrical stimulation was delivered only to the demonstrator, using Biopac STM100C and STMISOC modules, and controlled using a USB-6001 analogue output card (National Instruments, Texas, USA). Two electrodes with Ag/AgCl laminated, carbon composition contact (11 mm diameter; EL509) and cavities filled with salt-free conductive gel (Spectra 360, Parker Laboratories Inc.) were placed on the upper ventral part of the right forearm.

Data analysis

Behavioural results

Summary scores from questionnaires or subscales that were completed once (STAI-trait, BES-A) were compared using t-tests. Results of the STAI-state questionnaire, which was completed before and after the experiment, were analysed using a mixed-design ANOVA (with type 3 sums of squares) with measurement (before, after) and contingency knowledge (known, not known) as within- and between-subject factors, respectively. Individual items from the evaluation of the demonstrator and the observational US questionnaire were compared between groups using the Wilcoxon-Mann-Whitney test.

Electrodermal activity

The EDA signal was decomposed into tonic and phasic components using cvxEDA (Greco et al., 2016). Skin conductance responses were scored in the phasic component by taking the difference between the maximum value within 6 seconds after the stimulus onset and the mean value from the preceding 2 seconds. This method is analogous to the entire interval response approach (Pineles et al., 2009), except the response window duration was determined by startle probe timing rather than the US timing to avoid confounding CS responses by the startle probes. While the phasic signal should have a zero baseline by definition, baseline subtraction was performed to avoid scoring spontaneous fluctuations beginning before the stimulus. Amplitudes smaller than 0.2 µS were treated as no response. Responses to the observational US were analysed only for trials without startle probes. The "no US" responses were computed from non-reinforced CS+ trials, using the same time window as the US responses. Each subject's responses were normalised by calculating $\log(1 + \mathrm{SCR}/\mathrm{SCR}_{\mathrm{max}})$, where $\mathrm{SCR}_{\mathrm{max}}$ was the highest response observed for a given subject.

Three subjects with less than five non-zero responses during the direct-expression stage were excluded from the EDA analysis. All responses, including those with zero

amplitude, were averaged within conditions for each of the remaining subjects. The resulting mean magnitudes were used in group analysis.

A mixed-design ANOVA was computed separately for observational learning and direct-expression stages. The greenhouse-Geiser correction was applied if sphericity assumptions were broken. Group (known, unknown) was the between-subject factor, and stimulus (observational learning: CS+, CS–, US, no US; direct-expression: CS+, CS–) was the within-subject factor. Pairwise comparisons for each simple effect followed the ANOVA, and Holm correction was used.

Fear potentiated startle

The EMG signal was band-pass filtered (28–500 Hz, fourth-order Butterworth filter), rectified (converted to its absolute value), and smoothed using a 40 Hz low pass FIR filter following the guidelines (Blumenthal et al., 2005). Startle responses were scored by taking the difference between the peak value, registered 20–120 ms after stimulus onset, and the mean value from the 100 ms preceding the stimulus onset (Blumenthal et al., 2005). A trial was scored as 0 if the peak was below the baseline. The resulting amplitudes were then normalised for each subject by transforming them to T-scores.

All trials were visually inspected for the presence of artefacts. One participant was excluded from FPS analysis because the entire recording was noisy; other than that, no trials were marked for exclusion. Two further participants were excluded because they had less than five non-zero responses during the direct-expression stage. All responses, including those with zero amplitude, were averaged within conditions for each of the remaining subjects. The resulting mean magnitudes were used in group analysis.

Statistical analysis followed the same approach that was used for the electrodermal activity. A mixed-design ANOVA was computed separately for observational learning and direct-expression stages. The greenhouse-Geiser correction was applied if sphericity assumptions were broken. Group (known, unknown) was the between-subject factor, and condition (CS+, CS–, ITI) was the within-subject factor. Pairwise comparisons for each simple effect followed the ANOVA, and Holm correction was used.

Software

Custom scripts, written in Python and relying on NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), and bioread (Vack et al., 2019) libraries, were used for processing and analysis of the psychophysiological signals. Statistical analysis was performed in R, using the afex (Singmann et al., 2021), emmeans (Lenth, 2021), and rstatix (Kassambara, 2021) packages. The ggpubr library (Kassambara, 2020) was used to produce plots. In all ANOVAs, type 3 errors were used.

Table 2.1: Age, friendship length and friendship questionnaire scores of each participant subgroup. Only those demonstrators for whom either friend or stranger knew the contingency were included in age calculation. SD – standard deviation, IQR – interquartile range, obs – observer, dem – demonstrator, MFQ – McGill Friendship Questionnaire.

Variable	Subgroup	Mean (SD)	Median (IQR)	Range
Age	friend, obs	22.5 (2.77)	22 (4.5)	18–29
	stranger, obs	23.3 (2.95)	22.5 (4)	18–30
	demonstrators	22.1 (2.53)	22 (4)	18–28
Friendship length (years)	friend, obs	7.7 (4.22)	6 (7)	3–19
MFQ result	friend, obs	52 (7.74)	55 (9.5)	33–60
	friend, dem	49.8 (9.21)	52.5 (15)	30–60

2.3 Neuroimaging experiment

Participants

Forty-eight pairs of friends (friend group) and 47 individual participants (stranger group) participated in the experiment. All participants were male. In the friend group, one participant was randomly selected as the demonstrator, and the other was the observer. In the stranger groups, all participants were observers and watched recordings of the demonstrator from the friend group. Only the observers underwent fMRI.

To be eligible for the study, participants had to be heterosexual, aged between 18 and 30 years, and right-handed. They could not have neurological disorders or other medical conditions precluding MR scanning (such as having medical implants or stimulation devices) or electrical stimulation (such as heart diseases) and could not be using psychoactive medications. In the friend group, both participants had to meet the same criteria because roles were assigned after arrival to the laboratory. Students and graduates of either psychology or cognitive science were explicitly excluded at the recruitment stage, as their potential familiarity with fear conditioning procedures might have rendered the experimental manipulation ineffective. Additionally, pairs of friends had to have known each other for at least three years and score at least 30 out of 60 points (M = 50.8, SD = 8.7, range [30, 60]) in the McGill Friendship Questionnaire (Mendelson & Aboud, 1999). Most participants were undergraduate students from various academic institutions in Warsaw. All participants received financial remuneration of 100 PLN (~25 EUR) for their participation.

Four subjects (all from the stranger group) were excluded from the analysis for technical reasons: three due to video playback issues and one due to excessive head motion during fMRI. Moreover, only contingency-knowing participants were included in the analysis, which led to the final sample sizes of n = 34 (friends) and n = 34 (strangers). In the resulting sample, the mean age of all observers was 22.9 years (SD = 2.87). A detailed description is included in Table 2.1.

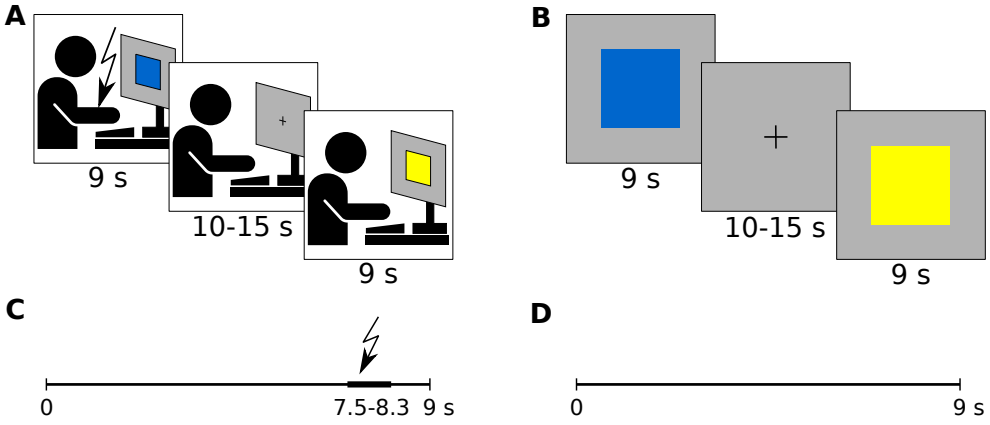


Figure 2.2: Trial sequence and within-trial timings for observational learning and direct-expression stages (as seen by the observers). In this example, the blue square is the CS+, and the yellow square is the CS–. (A) Trial sequence in the observational learning stage (half of CS+ were reinforced). (B) Trial sequence in the direct-expression stage. (C) During observational learning the electrical stimulation could be applied to the demonstrator in a series of five pulses, equally spaced from 7.5 to 8.3 s. (D) The electrical stimulation was never applied during direct-expression.

Procedure and tasks

The experiment was carried out according to the procedure described in the Shared methods section. A schematic depiction of the procedure is shown in Figure 2.2. Stimulus presentation was controlled using Presentation v 20.1 (Neurobehavioral Systems, Berkeley, CA, USA). During observational learning and direct-expression stages, the observer was lying in the MRI scanner. Apart from the task-related imaging, an anatomical scan lasting ca. 7 minutes had to be performed. This scan was collected before the tasks and allowed participants time to habituate to the environment. The time was also used for the demonstrator's preparations, not seen by the observers. In the friend group, the live-streamed videos of the demonstrators were also recorded, and they were later presented to observers from the stranger group. Every recorded demonstrator was seen by their friend and no more than one stranger[3].

In the stranger group, the observers arrived at the laboratory alone. They followed the same procedure as they would in the friend group, except the role assignment through envelope drawing was skipped (there was only one role), and instructions said they would watch another person (rather than their friend).

A GoPro Hero7 camera was used to transmit and record video. For transmission, an HDMI connection was used; the sound was not transmitted. A small room adjacent to the MR room was used for seating the demonstrator (see Figure 2.3). The room was lit

[3]Observers excluded from analysis due to playback issues are not counted here. The demonstrators were asked to provide optional consent for their recording to be displayed to other participants in the experiment. 45 out of 48 demonstrators agreed to reuse their recording.

(a) Outside view

(b) Demonstrator's room

Figure 2.3: The experimental setup. The demonstrator was sitting in the room adjacent to the MR room. Above the door, cables for the camera, stimulator and computer screen can be seen. The picture of the demonstrator's room shows what was visible to the camera (transmitted to the observer).

with an LCD panel, and the camera's white balance setting was manually adjusted to allow reliable reproduction of colours. Additionally, the brightness of the demonstrator's screen was set to low. Notably, the room's walls, table, and chair were either dark grey or black, which minimised visual distractors for the observer.

MRI data acquisition and preprocessing

Acquisition

Magnetic resonance imaging data were acquired using a 3 T Siemens Magnetom Trio scanner with a 12 channel head coil. At the beginning of a session, a T1-weighted anatomical image was acquired using an MPRAGE sequence with 1 × 1 × 1 mm resolution and the following parameters: inversion time (TI): 1100 ms, acquisition time (TA): 6 minutes and 3 seconds, GRAPPA parallel imaging with acceleration factor (PE): 2.

Two runs of functional imaging followed anatomical scans. The first run contained 362 volumes in the friend group (TA: 17 min 29 s) and 380 in the stranger group (TA: 18 min 21 s). The second run contained 184 volumes (TA: 8 min 58 s). The first run was longer in the stranger group because it simplified matching video presentation to scanning. The added volumes were removed before analysis to match data length between groups.

Each functional image volume comprised 47 axial slices (2.3 mm thick, with 2.3 × 2.3 mm in-plane resolution and 30% distance factor) that were acquired using a T2*-sensitive gradient echo-planar imaging (EPI) sequence with the following parameters: repetition time (TR): 2870 ms, echo time (TE): 30 ms, flip angle (FA): 90 degrees, field of view (FoV): 212 mm, matrix size: 92 × 92, with interleaved acquisition order, and using GRAPPA acceleration with factor 2.

Preprocessing - overview

The fMRI data preprocessing was performed using fMRIPrep 1.4.0 (Esteban, Blair et al., 2019; Esteban, Markiewicz et al., 2019), which is based on Nipype 1.2.0 (Gorgolewski et al., 2011; Gorgolewski et al., 2019). The fMRIPrep preprocessing was followed by smoothing using SPM 12 (v7487). In summary, after coregistration of anatomical and functional images, the functional images were motion corrected, slice-time corrected, normalised to the MNI template, resampled to 2 × 2 × 2 mm resolution, and smoothed with a 6 mm full width at half maximum Gaussian kernel. The preprocessing pipeline is described in detail below, based on the template generated by fMRIPrep.

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al., 2010), distributed with ANTs 2.2.0 (Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as the target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM) and grey-matter (GM) was performed on the brain-extracted T1w using `fast` (FSL 5.0.9; Zhang et al., 2001). Volume-based spatial normalisation to the standard MNI space was performed through nonlinear registration with `antsRegistration` (ANTs 2.2.0), using brain-extracted versions of the T1w reference and the T1w template. The following template was selected for spatial normalisation: ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009); TemplateFlow ID: MNI152NLin2009cAsym.

Functional data preprocessing

For each of the 2 BOLD runs found per subject, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The BOLD reference was then co-registered to the T1w reference using `flirt` (FSL 5.0.9; Jenkinson & Smith, 2001) with the boundary-based registration (Greve & Fischl, 2009) cost-function. Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using `mcflirt` (FSL 5.0.9; Jenkinson et al., 2002). BOLD runs were slice-time corrected using `3dTshift` from AFNI 20160207 (Cox & Hyde, 1997). The BOLD time-series were resampled onto their original, native space by applying a single composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD time-series were resampled into the standard space, generating a preprocessed BOLD run in the MNI space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Several confounding time-series were calculated based on

the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, using their implementations in Nipype (Power et al., 2014). The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file.

Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimise the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.5.2 (Abraham et al., 2014), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation (https://fmriprep.readthedocs.io/en/latest/workflows.html).

Following the fMRIPrep preprocessing, the functional images were smoothed with a 3-dimensional Gaussian kernel, 6 mm full width at half maximum, using the `spm_smooth` function from SPM 12.

Physiological recordings and stimulation

During fMRI scanning, skin conductance and pulse oximetry were registered as secondary measures. Skin conductance was recorded using BrainVision BrainAmp ExG MR amplifier with GSR MR sensor, sampled at 250 Hz. The electrodes were placed on distal phalanges of the right index and middle fingers. Pulse oximetry was recorded using a photoplethysmograph from the integrated Siemens Physiological Monitoring unit, sampled at 50 Hz, placed on the ring finger of the same hand.

Electrical stimulation consisting of five unipolar pulses (1 ms duration, 200 ms interval, resulting in total stimulation duration of 0.8 s) was used as the aversive stimulus. It was delivered using Biopac STM100C and STMISOC stimulators triggered using National Instruments USB-6001 analogue output card. Two Ag/AgCl laminated carbon composition contact electrodes placed 3.5 cm apart (measured between centres) and filled with salt-free electrode gel were used. They were placed on the upper ventral part of the right forearm. Mock electrode cables were used for participants in the scanner.

Behavioural data analysis

Summary scores from questionnaires or subscales that were completed once (STAI-trait, SWE) were compared using t-tests. Results of the STAI-state questionnaire, which was completed before and after the experiment, were analysed using a mixed-design ANOVA (with type 3 sums of squares) with measurement (before, after) and group (friend, stranger) as within- and between-subject factors, respectively. Individual items from the evaluation of the demonstrator and the observational US questionnaire were compared

between groups using the Wilcoxon-Mann-Whitney test. R (R Core Team, 2020) with afex (Singmann et al., 2021), emmeans (Lenth, 2021), and ggpubr (Kassambara, 2020) packages, was used.

Physiological data analysis

Skin conductance data were analysed using PsPM v4.3.0 software (https://bachlab.github.io/PsPM/) operating under Matlab 2018b. The non-linear model (Bach, Daunizeau et al., 2010) was chosen because it is well suited to fear conditioning data, where anticipatory responses typically occur with an unknown and variable latency after stimulus onset. For each trial in observational learning, two responses were modelled: a flexible (allowed latency 0–7.5 s) response to the CS presentation and a fixed response to the US presentation. In direct-expression, only the response to the CS presentation was modelled (allowed latency 0–7.5 s), as the US were not delivered in this stage. Default PsPM settings were used, including high- and low-pass filter cut off frequencies of 0.0159 and 5 Hz, respectively, resampling to 10 Hz and inversion of 2 trials at the same time (Staib et al., 2015). GNU Parallel software (Tange, 2011) was used to accelerate the processing of multiple subjects.

Before analysis, the recorded signals were visually screened for the presence of artefacts. Periods containing artefacts, most often caused by electrodes temporarily losing contact, were marked for exclusion from data scoring. Due to technical problems with data recording or an excessive presence of artefacts that precluded further processing, the SCR analysis included data from 27 subjects per group.

Amplitude parameters obtained for each trial were analysed in R (R Core Team, 2020) with methods implemented in afex (Singmann et al., 2021) and emmeans (Lenth, 2021) packages. A mixed-design ANOVA was computed separately for observational learning and direct-expression stages. The greenhouse-Geiser correction was applied if sphericity assumptions were broken. Group (friend, stranger) was the between-subject factor, and stimulus (observational learning: US, no US; direct-expression: CS+, CS–) was the within-subject factor. Pairwise comparisons for each simple effect followed the ANOVA, and Holm correction was used. In addition to the classical ANOVA, an analogous Bayesian ANOVA was computed using JASP (JASP Team, 2020) with default priors, and effects are reported as the Bayes factor for the inclusion of a given effect, calculated as the ratio between the likelihood of the data given all the models with vs without the effect.

fMRI data analysis

Region of interest specification

Regions of interest (ROIs) were defined to be used in ROI and PPI analyses. In an ROI analysis, voxelwise activation estimates are averaged within a given ROI, while in a PPI analysis, the ROI is used to extract BOLD time series. Anatomical definitions were selected based on a meta-analysis focusing on social processing (Alcalá-López et al., 2018) and its related Neurovault dataset (https://identifiers.org/neurovault.collection:

2462). The following six regions were selected: amygdala, anterior insula (AI), anterior mid-cingulate cortex (aMCC), fusiform face area (FFA), right temporoparietal junction (rTPJ), and right posterior STS (rpSTS). Bilateral masks of the amygdala, AI, and FFA were created by merging their hemispheric definitions; the aMCC is located centrally and already had a single mask in the collection. Conversely, masks from the right hemisphere were used for rTPJ and rpSTS because social processing in these structures is considered to be primarily right-lateralized (Alcalá-López et al., 2018; Boccadoro et al., 2019; Sliwinska & Pitcher, 2018).

Additionally, a separate definition of the amygdala ROI was used for small volume correction (SVC) for multiple comparisons. This ROI was more extensive and was based on the Harvard-Oxford atlas thresholded at 0.7, following (Lindström et al., 2018). The SVC within the amygdala was performed to complement the whole-brain corrections.

Model specification

The fMRI data were analysed using a mass univariate approach based on the general linear model in SPM 12 (v7484) running under Matlab R2020a. In the observational learning stage, the following four events were modelled: CS+ onset, CS– onset, US occurrence (further: US), US omission (further: no US). The CS onsets were modelled as instantaneous events (0 s duration), while the US and no US were modelled with 1.5 s duration (from the potential US onset to the end of CS). The no US event was included for nonreinforced CS+ trials, and its within-trial timing was the same as the timing of the US. In the direct-expression stage, two events were modelled: CS+ and CS–, both with a 9 s duration. Additionally, linear temporal modulation was included to capture extinction effects during this stage (two added regressors: CS+ × time, CS– × time). While (Lindström et al., 2018) modelled all events as instantaneous, here, nonzero durations were chosen for the events upon which activation and PPI analyses were based.

Additional regressors: six estimated head motion parameters (three rotations and three translations) and a variable number of motion spike regressors generated by fMRIPrep were also included in the models for observational learning and direct-expression. A motion spike regressor has a value of one for a given outlier volume and zeroes for all others. There were on average 13.4 motion spike regressors in observational learning (Mdn: 8, range 0–60) and 5.5 in direct-expression (Mdn 2, range 0–38).

Contrast definition and whole-brain activation analysis

In the primary analysis aiming to evaluate acquisition and expression of fear, two contrasts were evaluated: US > noUS (observational learning) and CS+ > CS– (direct-expression). Contrast estimates obtained for each subject were entered into the second-level analysis. First, both groups were analysed together to test the main effects of the tasks (one-sample t-test) and then the groups were compared (two-sample t-test). The resulting statistical maps were thresholded at $p < 0.05$ using a peak-level FWE correction.

An additional analysis was performed, in which the CS+ × time parameter estimates with a reversed sign (therefore representing a linear decrease of response in time) were

entered into the second-level analysis. As previously, a one-sample t-test with all subjects was evaluated first, and followed by a 2 sample t-test comparing between groups. The resulting statistical maps were thresholded at $p < 0.05$ using a cluster-level FWE correction with a cluster defining threshold of $p = 0.001$.

Region of interest analysis

A region of interest analysis was performed for the US > noUS (observational learning) and CS+ > CS– (direct-expression) to further investigate the effects tested in the whole-brain analysis. Parameter estimates were averaged within each ROI defined above, producing a single estimate per region for each subject. The estimates were compared between groups using t-tests and Bayesian t-tests using the BayesFactor (Morey & Rouder, 2018) R package.

PPI analysis

Finally, a PPI analysis was performed using the ROIs outlined above as seeds. An identical procedure was repeated for each seed. The analysis was based on the US > noUS contrast (psychological variable) in observational learning and CS+ > CS– in the direct-expression. The first eigenvariate of the motion-regressed BOLD signal was extracted for each region and used as a physiological variable. The psychophysiological interaction term was calculated using SPM's methodology and entered into the PPI design matrix, together with the physiological and psychological variables. Head motion parameters and motion spike regressors were also included. Parameter estimates obtained for the interaction term were entered into the second-level analysis. Shared effects were tested with a 1 sample t-test, and groups were compared using a 2 sample t-test. Obtained statistical maps were thresholded at $p < 0.05$ using a cluster-level FWE correction with a cluster defining threshold of $p = 0.001$.

Chapter 3

Results

3.1 Psychophysiological experiment

Behavioural results

Contingency knowledge

In the questionnaire referring to the observational learning stage (completed at the end of the experiment), 14 participants correctly identified the contingency of CS+ and US, and 21 participants incorrectly. Consequently, participants were divided into two groups for further analyses.

Empathy and trait anxiety questionnaires

The STAI-trait and the three subscales of BES (cognitive empathy, emotional contagion, emotional disconnection) were treated independently and compared between contingency-knowing and not-knowing participants using t-tests. The groups did not differ significantly in either of the measures. Summary statistics and comparison results are presented in Table 3.1, and the results are plotted in Figure 3.1.

Change in the state anxiety during the experiment

The STAI-state scores, collected at the beginning and the end of the experiment, were entered into the analysis of variance, with measurement as a within-subject factor and contingency as a between-subject factor. There was a significant main effect of measurement, $F(1, 33) = 6.67$, $\eta_g^2 = .032$, $p = .014$. The main effect of contingency was not significant, $F(1, 33) = 0.46$, $\eta_g^2 = .012$, $p = .501$, and neither was the contingency × measurement interaction, $F(1, 33) = 0.24$, $\eta_g^2 = .001$, $p = .631$. At the end of the experiment, the STAI-state score was, on average, 2.35 points (SEM = 0.91) lower than at the beginning. Individual results are plotted in the Supplementary Figure 1 A.

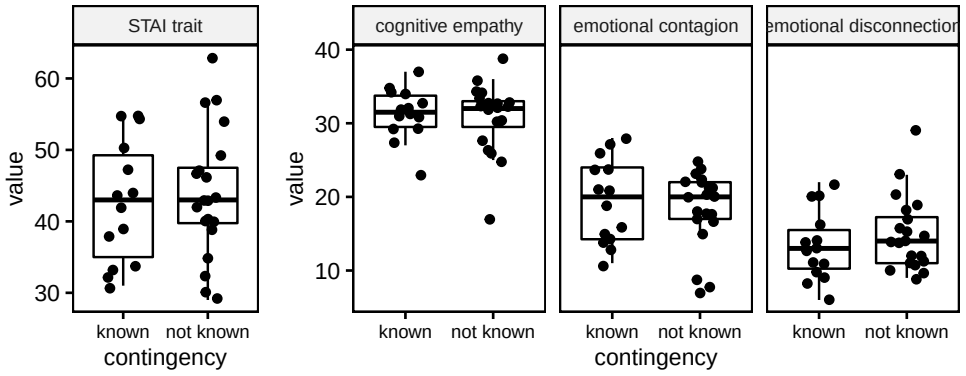


Figure 3.1: Results of STAI-trait and the three subscales of BES-A questionnaires. Dots represent individual subjects. The lower and upper hinges correspond to the first and third quartiles, and whiskers extend to the smallest (lower) or largest (upper) value no further than 1.5 × IQR from the hinge. K – known, NK – not known.

Table 3.1: Results of STAI-trait and BES-A questionnaires split by contingency knowledge. The comparison column lists the uncorrected t-test results. Mdn – median, IQR – interquartile range.

Scale	Contingency	Mdn	IQR	Range	Comparison
STAI					
Trait anxiety	known	43	14.2	31–55	$t(29.1) = -0.35$
	not known	43	7.75	29–63	$p = .73$
BES-A					
Cognitive empathy	known	31.5	4.75	27–37	$t(31.8) = 0.38$
	not known	32	3.5	17–39	$p = .71$
Emotional contagion	known	20	9.75	11–28	$t(26.1) = 0.60$
	not known	20	5	7–25	$p = .55$
Emotional disconnection	known	13	5.25	6–22	$t(28.9) = -0.97$
	not known	14	6.25	9–29	$p = .34$

Evaluation of the demonstrator and the observational US

The observers evaluated the behaviour of the demonstrator (discomfort, expressiveness and naturalness when reacting to the aversive stimulation) and their own reactions (empathy toward the demonstrator and identifying with them) after the experiment. The ratings, given for each question using a discrete 0–9 range, were compared between groups using the Wilcoxon-Mann-Whitney test, and no statistically significant differences were found. Results of the ratings are shown in Figure 3.2; summary statistics and comparison results are presented in Table 3.2.

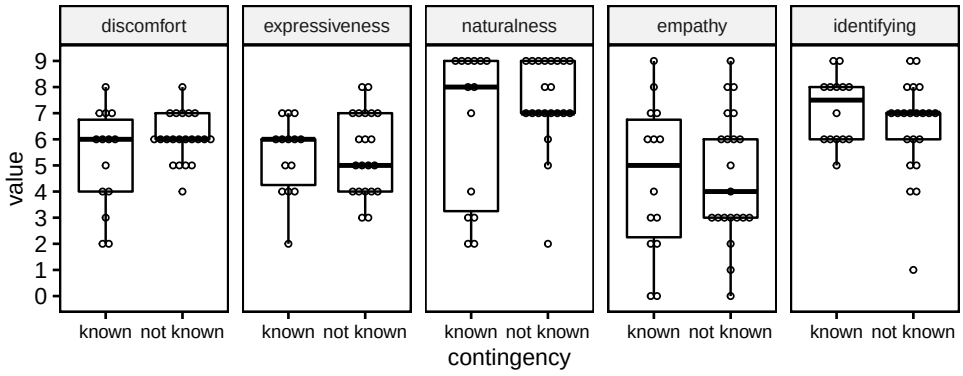


Figure 3.2: Evaluation of the demonstrator and the observational US. Using five discrete scales, the observers evaluated the demonstrator's reactions to the aversive stimulation (first three panels) and their own attitude towards the demonstrator (last two panels). Dots represent individual subjects. The lower and upper hinges correspond to the first and third quartiles, and whiskers extend to the smallest (lower) or largest (upper) value no further than 1.5 × IQR from the hinge.

Table 3.2: Evaluation of the demonstrator and the observational US, split by contingency knowledge. The available response range was between 0 and 9. The comparison column lists the results of the Wilcoxon-Mann-Whitney test without corrections for multiple comparisons. Mdn – median, IQR – interquartile range.

Rating	Contingency	Mdn	IQR	Range	Comparison
discomfort	known	6	2.75	2–8	$W = 117, p = .30$
	not known	6	1	4–8	
expressiveness	known	6	1.75	2–7	$W = 144, p = .92$
	not known	5	3	3–8	
naturalness	known	8	5.75	2–9	$W = 137, p = .74$
	not known	7	2	2–9	
empathy	known	5	4.5	0–9	$W = 145, p = .96$
	not known	4	3	0–9	
identifying	known	7.5	2	5–9	$W = 176, p = .33$
	not known	7	1	1–9	

Physiological results

Analysis of variance was carried out separately for observational learning and direct-expression stages, with stimulus as a within-subject factor and contingency knowledge as a between-subject factor. The analyses for EDA and FPS had an analogous structure, although the levels of stimulus factor were different. For EDA, there were four factor levels for observational learning: CS+, CS–, US, noUS, and two factor levels for direct-

expression: CS+ and CS–. For FPS, there were three levels for each stage: CS+, CS– and ITI. The ANOVA was followed up by post hoc pairwise comparisons for all simple effects, and Holm correction for multiple comparisons was used.

Electrodermal activity

In observational learning there was a significant effect of stimulus, $F(1.75, 52.39) = 47.15$, $\eta_g^2 = .409$, $p < .001$, while both the effect of contingency, $F(1, 30) = 0.17$, $\eta_g^2 = .003$, $p = .68$, and the contingency × stimulus interaction, $F(1.75, 52.39) = 0.71$, $\eta_g^2 = .010$, $p = .48$ were not significant. Post-hoc comparisons confirmed that the responses to the US were significantly higher than for all the other stimuli. Specifically, in the most relevant US - no US comparison, $t(90) = 7.96$, $p < .001$ for contingency-knowing and $t(90) = 6.84$, $p < .001$ for not-knowing participants (Figure 3.3 A).

In direct-expression, there was a significant contingency × stimulus interaction, $F(1, 30) = 12.29$, $\eta_g^2 = .099$, $p = .001$. The main effects of contingency, $F(1, 30) = 1.16$, $\eta_g^2 = .028$, $p = .29$ and stimulus, $F(1, 30) = 2.24$, $\eta_g^2 = .020$, $p = .15$ were not significant. Post hoc comparisons identified two significant effects. First, reactions to CS+ were greater than to CS– for the contingency-knowing participants, $t(30) = 3.34$, $p = .01$. Second, the contingency knowing participants had higher reactions to CS+ than the not knowing participants, $t(49.3) = 2,73$, $p = .03$ (Figure 3.3 B).

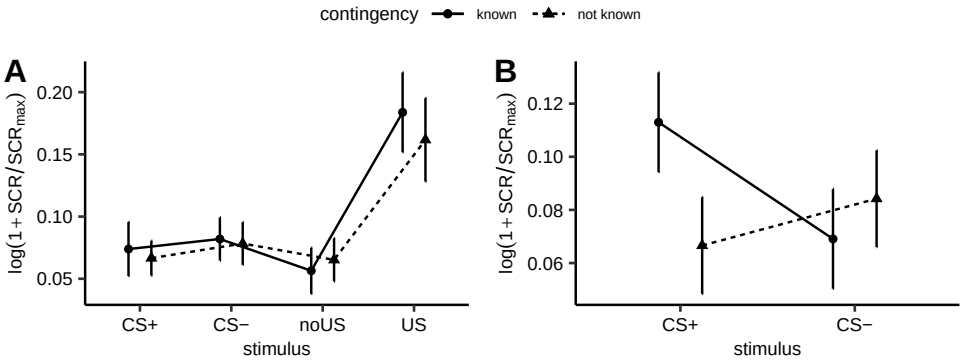


Figure 3.3: Skin conductance responses. Points show the response means for each group and condition included in ANOVA. Error bars show 95% confidence intervals based on within-subject standard errors estimated using the Cousineau-Morey-O'Brien method. (A) Observational learning stage. The main effect of the stimulus was statistically significant. In both groups, responses to the US were significantly higher than to all the other stimuli. There were no statistically significant between-group differences. (B) Direct-expression stage. The contingency × stimulus interaction was statistically significant. Reactions to CS+ in the contingency knowing group were higher than reactions to CS– in the same group and higher than reactions to CS+ in the group not knowing the contingency.

Fear potentiated startle

In observational learning there was a significant main effect of stimulus, $F(1.94, 58.08) = 6.37$, $\eta_g^2 = .124$, $p = .003$. The main effect of contingency, $F(1, 30) = 0.03$, $\eta_g^2 < .001$, $p = .87$, and the contingency × stimulus interaction, $F(1.94, 58.08) = 0.11$, $\eta_g^2 = .003$, $p = .89$, were not significant. None of the simple contrasts was significant when accounting for multiple comparisons. However, when averaging over the levels of contingency, responses during both CS+ and CS− were higher than during ITI, $t(60) = 3.02$, $p = 0.008$ and $t(60) = 3.16$, $p = 0.008$ respectively (see Figure 3.4 A).

In the direct-expression stage there was a significant main effect of stimulus, $F(1.82, 54.71) = 8.15$, $\eta_g^2 = .154$, $p = .001$, and a trend-level contingency × stimulus interaction, $F(1.82, 54.71) = 2.59$, $\eta_g^2 = .055$, $p = .089$. The main effect of contingency was not significant, $F(1, 30) = 0.03$, $\eta_g^2 < .001$, $p = .87$. Post hoc comparisons revealed only one significant difference, namely the reactions during CS+ were higher than during ITI for contingency knowing participants, $t(60) = 3.96$, $p = .0018$. Additionally, there was a trend-level difference of CS− vs ITI also for the contingency-knowing participants, $t(60) = 2.78$, $p = .058$. The primarily interesting comparison of CS+ vs CS− was not significant in both groups, $t(60) = 1.18$, $p = .91$ and $t(60) = -0.16$, $p = 1.00$ for contingency knowing and not-knowing participants respectively (see Figure 3.4 B).

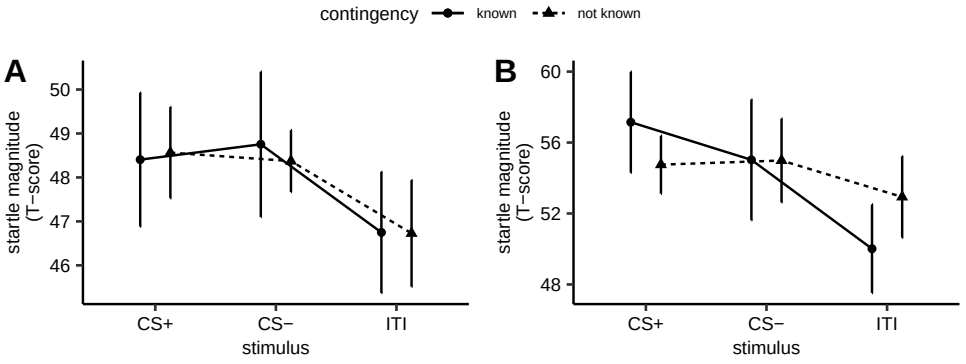


Figure 3.4: Fear potentiated startle responses. Points show the response means for each group and condition included in ANOVA. Error bars show 95% confidence intervals based on within-subject standard errors estimated using the Cousineau-Morey-O'Brien method. (A) Observational learning stage. The main effect of the stimulus was statistically significant. Responses, averaged over the levels of contingency, were higher during both CS+ and CS− than during ITI. (B) Direct-expression stage. The main effect of the stimulus was statistically significant, and there was a trend towards a contingency × stimulus interaction. For contingency-knowing participants, reactions during CS+ were higher than during ITI, and there was a trend-level difference between CS− and ITI.

3.2 Neuroimaging experiment

Behavioural results

Contingency knowledge

The stimulus contingency questionnaire was completed at the end of the experiment and tested whether observers associated the CS+ with the US during observation. 35 out of 48 participants in the friend group and 35 out of 44 in the stranger group correctly identified the stimulus contingency. The proportion of contingency-knowing participants was not significantly different between the friend and stranger groups, $\chi^2 = 0.554$, $df = 1$, $p = 0.46$. Considering that contingency knowledge was a prerequisite for fear conditioning in the previous experiment, only the contingency-knowing participants were included in further analyses.

Empathy and anxiety questionnaires

The STAI-trait and SWE questionnaires were completed at the end of the experiment and were used to evaluate participants' anxiety and empathy. The STAI-trait and three subscales of SWE (empathic concern, personal distress, perspective-taking) were treated independently and compared between groups using t-tests. The groups did not differ significantly in any of the measures. Summary statistics and comparison results are listed in Table 3.3 and presented in Figure 3.5.

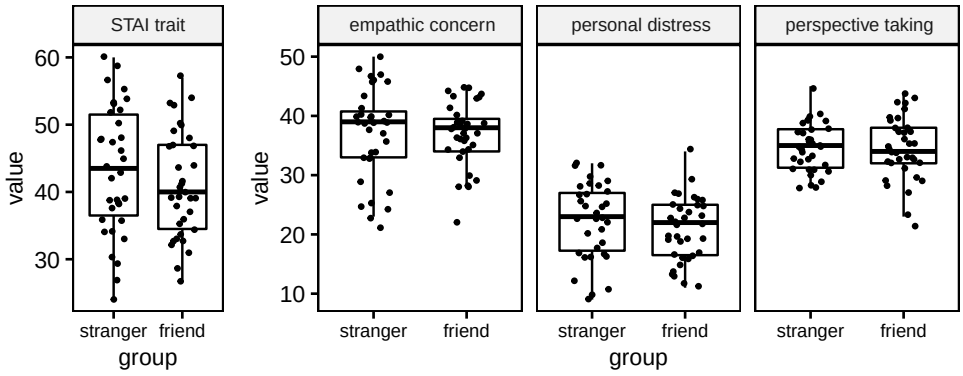


Figure 3.5: Results of STAI-trait and the three subscales of SWE questionnaires. Dots represent individual subjects. The lower and upper hinges correspond to the first and third quartiles, and whiskers extend to the smallest (lower) or largest (upper) value no further than 1.5 × IQR from the hinge.

Change in state anxiety during the experiment

Observers completed the STAI-state questionnaire twice, before and after the experiment. The obtained scores were entered into analysis of variance, with group (friend,

Table 3.3: Results of anxiety and empathy questionnaires split by the group. The comparison column lists the uncorrected t-test results. Mdn – median, IQR – interquartile range.

Scale	Group	Mdn	IQR	Range	Comparison
STAI					
trait anxiety	friend	40	12.5	27–57	$t(63.5) = -1.07$,
	stranger	43.5	15	24–60	$p = 0.29, \mathrm{BF}_{01} = 2.47$
SWE					
empathic concern	friend	38	5.5	22–45	$t(59.0) = -0.20$,
	stranger	39	7.75	21–50	$p = 0.85, \mathrm{BF}_{01} = 3.97$
personal distress	friend	22	8.5	11–34	$t(64.1) = -1.11$,
	stranger	23	9.75	9–32	$p = 0.27, \mathrm{BF}_{01} = 2.39$
perspective taking	friend	34	6	21–44	$t(63.2) = -0.12$,
	stranger	35	6.5	28–45	$p = 0.91, \mathrm{BF}_{01} = 4.02$

stranger) as a between-subject factor and measurement (before, after) as a within-subject factor. There were no statistically significant effects; main effect of group: $F(1, 67) = 0.76$, $\eta_g^2 = .009$, $p = .39$, $\mathrm{BF}_{\mathrm{incl}} = 0.35$; main effect of measurement: $F(1, 67) = 1.92$, $\eta_g^2 = .007$, $p = .17$, $\mathrm{BF}_{\mathrm{incl}} = 0.36$; group × measurement interaction $F(1, 67) = 1.29$, $\eta_g^2 = .004$, $p = .259$, $\mathrm{BF}_{\mathrm{incl}} = 0.14$ (see Supplementary Figure 1 B).

Evaluation of the demonstrator and the observational US

After the experiments, observers evaluated the demonstrator's responses to the aversive stimulation and their perception of these responses by answering five (friends) or four (strangers) questions on 0–9 discrete scales. Higher values indicated greater perceived strength, expressiveness or naturalness of reactions, greater empathy, and a greater degree of identification with the demonstrator (the last question was not asked in the stranger group). All median ratings in both groups were between 5 and 7, see Figure 3.6. Each rating was compared between groups using a Wilcoxon-Mann-Whitney test, and no significant differences were found (see Table 3.4).

Physiological results: electrodermal activity

The SCR amplitudes were analysed (separately for the two stages of the experiment) using classical and Bayesian repeated-measures ANOVA with the group (friend, stranger) as a between-subject factor. In the observational learning stage, ANOVA revealed a significant main effect of the stimulus (US / noUS; the observational CS were not included in the analysis), $F(1, 52) = 29.19$, $\eta_g^2 = .089$, $p < .001$, $\mathrm{BF}_{\mathrm{incl}} = 6976$. Responses were higher to the US compared to noUS. The main effect of the group, $F(1, 52) = 2.21$, $\eta_g^2 = .034$, $p =$

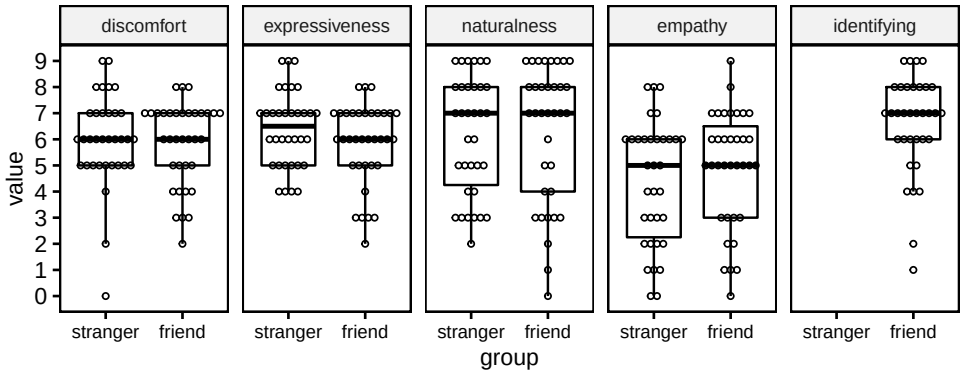


Figure 3.6: Evaluation of the demonstrator and the observational US. Using five discrete scales, the observers evaluated the demonstrator's reactions to the aversive stimulation (first three panels) and their own attitude towards the demonstrator (last two panels). Dots represent individual subjects. The lower and upper hinges correspond to the first and third quartiles, and whiskers extend to the smallest (lower) or largest (upper) value no further than 1.5 × IQR from the hinge.

Table 3.4: Evaluation of the demonstrator and the observational US, split by group. The available response range was between 0 and 9. The comparison column lists the results of the Wilcoxon-Mann-Whitney test without corrections for multiple comparisons. Mdn – median, IQR – interquartile range.

Rating	Group	Mdn	IQR	Range	Comparison
discomfort	friend	6	2	2–8	$W = 582,\ p = 0.87$
	stranger	6	2	0–9	
expressiveness	friend	6	2	2–8	$W = 501,\ p = 0.25$
	stranger	6.5	2	4–9	
naturalness	friend	7	4	0–9	$W = 630,\ p = 0.67$
	stranger	7	3.75	2–9	
empathy	friend	5	3.5	0–9	$W = 672,\ p = 0.36$
	stranger	5	3.75	0–8	
identifying	friend	7	2	1–9	n/a
	stranger	n/a	n/a	n/a	

.14, $BF_{incl} = 0.75$, and the stimulus × group interaction, $F(1, 52) = 0.55$, $\eta^2_g = .002$, $p = .46$, $BF_{incl} = 0.63$, were not significant (Figure 3.7 A).

Similarly, in the direct-expression stage, there was a significant main effect of the stimulus (CS+ / CS–), $F(1, 52) = 5.73$, $\eta^2_g = .018$, $p = .02$, $BF_{incl} = 1.79$. Responses were higher to CS+ than to CS–. The main effect of the group, $F(1, 52) = 1.89$, $\eta^2_g = .029$, $p = .18$, $BF_{incl} = 0.62$, and the stimulus × group interaction, $F(1, 52) = 0.83$, $\eta^2_g = .003$, $p = .37$, $BF_{incl} = 0.464$, were not significant (Figure 3.7 B).

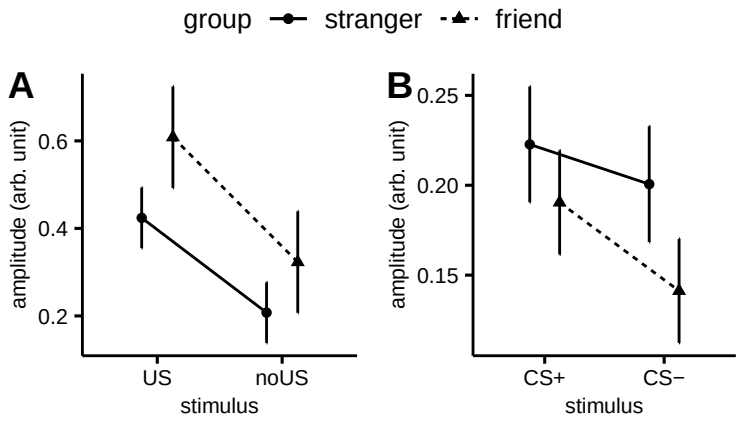


Figure 3.7: Skin conductance responses. Points show the response means for each group and condition included in ANOVA. Error bars show 95% confidence intervals based on within-subject standard errors estimated using the Cousineau-Morey-O'Brien method. Amplitudes are expressed in units of sudomotor nerve activity such that a standard-width pulse with unit amplitude would cause an evoked SCR with 1 µS amplitude. The main stimulus effects were significant, while group effects and stimulus × group interactions were not. (A) Observational learning stage. (B) Direct-expression stage.

Imaging results

Brain activation analysis - observational learning stage

Activation analysis in the observational learning stage focused on the perception of the observational US. Given the novelty of the paradigm, the analysis was carried out in two steps. First, subjects from both groups were analysed together to evaluate the main effect of the task. Next, the friend and stranger groups were compared directly.

The US > noUS contrast evaluated for both groups resulted in an extensive and robust activation of multiple brain areas (see Figure 3.8, Supplementary Table 2). Notably, subjects activated fear-relevant regions: the amygdala (bilateral), anterior mid-cingulate cortex, and anterior insula (bilateral). The insula activation extended to the operculum and orbitofrontal cortex (posterior orbital gyrus). Subjects also activated regions relevant to the perception of social stimuli, such as the STS (extending from its posterior part to the temporal pole in the right hemisphere, less extensive in the left) or the bilateral fusiform gyrus. Other notable regions included the inferior occipital gyrus (bilateral), the supplementary motor cortex (bilateral, more extensive in the right hemisphere), and the thalamus. The between-group comparison yielded no significant results.

A region of interest analysis (averaging parameter estimates within a region) was carried out for six preselected areas (AI, aMCC, amygdala, FFA, rpSTS, rTPJ), defined independently from the functional results, to further investigate the between-group effects. No significant differences were found between friend and stranger groups in any

of the regions. Bayes factors indicate moderate evidence for the absence of effects (BF_{01} > 3) in all regions except the rTPJ, where the evidence was inconclusive (BF_{01} = 2.07), see Figure 3.8 B and Table 3.5.

Additionally, the between-group comparisons were repeated based only on the US responses (i.e. friend US vs stranger US). There were no significant differences in the whole brain and region of interest analyses. Bayes factors again indicate moderate or inconclusive evidence for the absence of effects (BF_{01} between 1.36 and 4.04); see Supplementary Figure 2 and Supplementary Table 1.

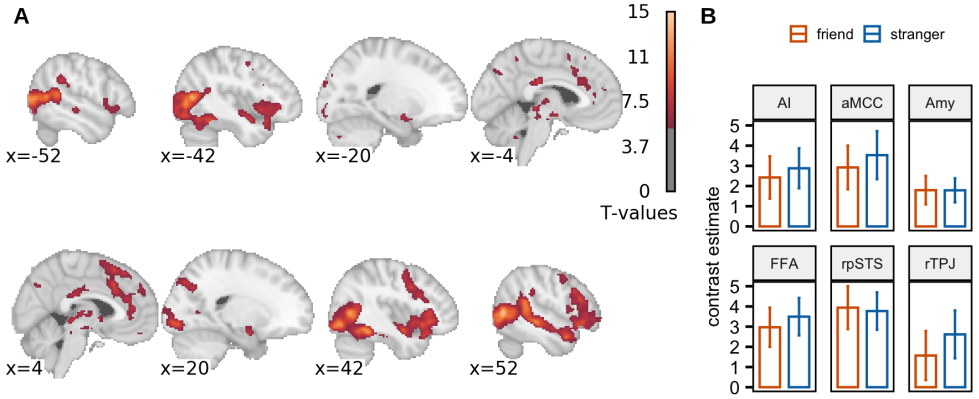


Figure 3.8: Activation in the observational learning stage, US > noUS contrast. The contrast revealed a widespread activation in both groups of subjects, with no significant differences between groups. (A) Cross-sections showing results for both groups analysed together, $p < 0.05$ FWE peak-level correction. (B) Parameter estimates and 95% confidence intervals for selected regions of interest. AI – anterior insula, aMCC – anterior mid-cingulate cortex, Amy – amygdala, FFA – fusiform face area, rpSTS – right posterior superior temporal sulcus, rTPJ – right temporoparietal junction.

Table 3.5: Region-of-interest analysis of between-group differences

	observational learning (US > noUS)			direct-expression (CS+ > CS–)		
ROI	t	p	BF_{01}	t	p	BF_{01}
AI	−0.64	.52	3.39	0.03	.97	4.04
aMCC	−0.77	.45	3.13	−0.08	.93	4.03
amygdala	0.01	.99	4.04	0.40	.68	3.76
FFA	−0.80	.43	3.08	−0.28	.78	3.90
rpSTS	0.24	.81	3.94	1.14	.26	2.32
rTPJ	−1.25	.21	2.07	−0.13	.90	4.01

Brain activation analysis - direct-expression stage

Activation analysis in the direct-expression stage focused primarily on the CS+ > CS– contrast, which reflects the fear-conditioned response. The analysis was first carried out for both groups together to reveal the shared effects, and followed by a between-group comparison. The CS+ > CS– contrast evaluated for both groups resulted in an activation of bilateral AI and aMCC (Figure 3.9, Supplementary Table 3). The between-group comparison yielded no significant differences.

Next, a region of interest analysis was carried out for the same areas as for the observational learning stage. No significant differences were found between groups. Bayes factors indicate moderate evidence for the absence of effects ($BF_{01} > 3$) in all regions except the rpSTS, where the evidence was inconclusive ($_{01} = 2.32$).

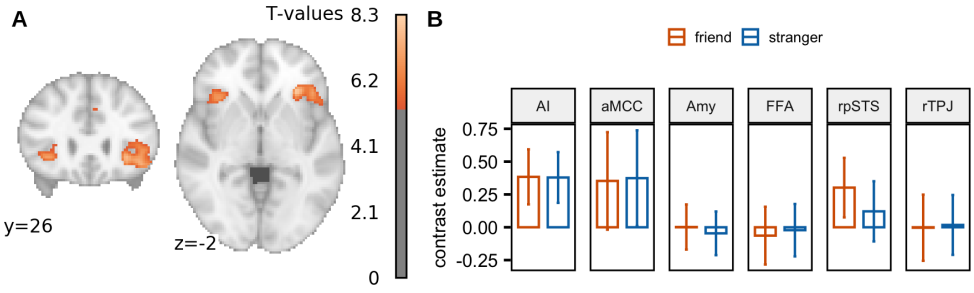


Figure 3.9: Activation in the direct-expression stage, CS+ > CS– contrast. Significant activations were found in the anterior insula and anterior mid-cingulate contrasts for both groups. There were no significant between-group differences. (A) Cross-sections showing results for both groups analysed together, $p < 0.05$ FWE peak-level correction. (B) Parameter estimates and 95% confidence intervals for selected regions of interest. AI – anterior insula, aMCC – anterior mid-cingulate cortex, Amy – amygdala, FFA – fusiform face area, rpSTS – right posterior superior temporal sulcus, rTPJ – right temporoparietal junction.

Finally, the effects of fear extinction were tested based on the CS+ temporal modulation regressor included in the GLM. For both groups analysed together, there was a significant decrease in the CS+ responses with time in several brain regions, including the aMCC (Figure 13 A, Table Y). For a between-group analysis reflecting differences in temporal dynamics of brain responses (friend CS+ × time > stranger CS+ × time), several significant clusters were identified. Strangers showed a more substantial decrease in responses in the left inferior frontal gyrus, left middle temporal gyrus, right lingual gyrus, and left putamen (Figure 3.10, Supplementary Table 4). Notably, the differences were not found in the areas activated in the observational and direct-expression stages.

Psychophysiological interactions

The psychophysiological interactions analysis was performed to investigate the coupling between the task-relevant structures. First, the observational learning stage was analysed

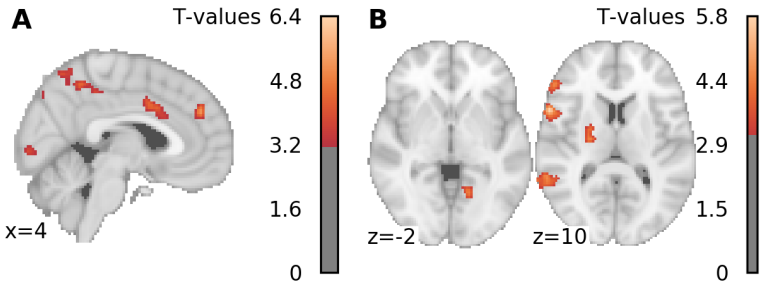


Figure 3.10: Temporal modulation of CS+ responses in the direct-expression stage. Cluster size correction was used, $p < 0.05$ FWE. (A) Clusters showing responses significantly decreasing in time for both groups analysed together. One of the clusters was located in the anterior mid-cingulate cortex. (B) Results of the (friend CS+ × time) > (stranger CS+ × time) contrast.

based on the US > noUS contrast with AI as the seed region, similar to (Lindström et al., 2018). Analysis for both groups showed several clusters where the PPI effect was present, including the rpSTS (Figure 3.11 A, Supplementary Table 5). However, the rpSTS cluster extended in the posterior direction and mainly covered the right lateral occipital cortex. Therefore, the PPI analysis was repeated with rpSTS (defined independently of the result) as the seed, confirming that the rpSTS increased its coupling with the AI, as well as the right fusiform gyrus and the amygdala (Figure 3.11 B, Supplementary Table 5)..

The between-group effects were tested for the above two seed regions (AI, rpSTS) and the aMCC, amygdala, FFA, and rTPJ. No significant differences were observed for any of the analyses. Further, the PPI analysis was applied to the direct-expression stage, using the AI, aMCC, and amygdala as the seeds. No significant PPI effects were found, neither for both groups analysed together nor between the groups.

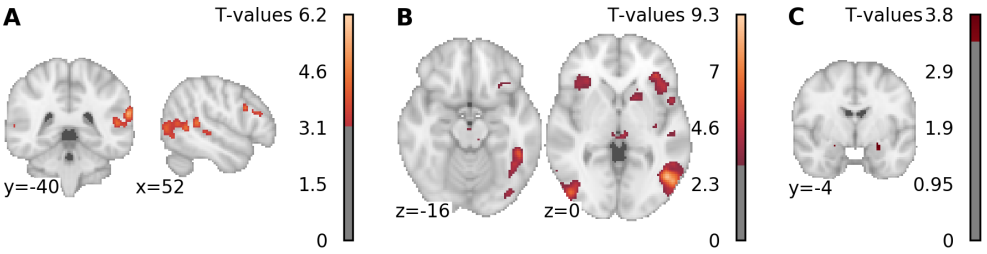


Figure 3.11: The psychophysiological interaction analysis results in the observational learning stage, based on the US > noUS contrast. (A) Clusters exhibiting an interaction effect with the amygdala seed. Cluster size correction, $p < 0.05$. (B) Clusters exhibiting an interaction effect with the rpSTS seed. Cluster size correction, $p < 0.05$. (C) Voxels in the amygdalae exhibiting an interaction effect with the rpSTS seed, $p < 0.05$ SVC.

Chapter 4

Discussion

The experiments described in this thesis investigated the mechanisms of observational fear conditioning in pairs of friends. The two experiments used a modified version of an existing paradigm, intending to increase naturalness and ecological validity by employing live observation. The psychophysiological experiment was conducted early in the research project to establish the validity of the experimental design. In addition to serving as a proof-of-concept for live observation, it was only the second observational fear conditioning experiment that used fear-potentiated startle, following (Selbing & Olsson, 2019) and the first to combine startle and skin conductance response measurements. Results of this experiment were previously published (Szczepanik et al., 2020), but here they are discussed in the broader context of the research project. The knowledge and procedural experience gained from the psychophysiological experiment were used in the neuroimaging study. This was the first observational fear conditioning experiment using real-time observation in fMRI and the first to investigate the effects of familiarity on observational fear conditioning.

The obtained results suggest that social transfer of fear between friends studied through live observation in a laboratory setting is effective (as evidenced by psychophysiological and behavioural measures). However, it is not enhanced in friends compared to strangers (an analogous group of observers without prior knowledge of the person they observed). Below, I will first discuss the results of the psychophysiological experiment, then the fMRI results, and in the end, I will provide a broader discussion of the two experiments.

4.1 Conclusions from the psychophysiological experiment

Declarative contingency knowledge

The first observation from the psychophysiological experiment was that the rate of declarative learning was unexpectedly low. In a questionnaire at the end of the experiment, less than half of the participants correctly indicated which stimulus predicted the occurrence of the US during observational learning. Crucially, declarative learn-

ing was reflected in physiological outcome measures, when contingency-knowing and not-knowing participants were treated as separate groups.

Several factors may have contributed to the low contingency knowledge ratio. First, the instructions given to participants informed them about the presentation of visual stimuli and possible administration of electrical stimulation but did not suggest any relationship between the two. A study directly investigating how stimulus contingency instructions influence conditioning outcomes has shown that contingency knowledge, SCR and FPS (Mertens et al., 2021) are all affected by the content of the instructions given to the participants[1]. Next, the reinforcement ratio was 50%, which weakened the learning situation. Additionally, startle probes may have served as distractors, and their inclusion has been shown to interfere with fear acquisition (Sjouwerman et al., 2016). Finally, the demonstrators were seated in a regular room, rather than against a uniform and contrasting background, which may have introduced additional visual distractors for the observer.

How does the observed contingency knowledge ratio (40%) compare to the neuroimaging experiment and experiments known from the literature? In the neuroimaging experiment, the ratio was higher (76%). The literature provides different reports. On the one hand, a similar fMRI study of observational fear conditioning (Lindström et al., 2018) reported no contingency unaware participants; however, that experiment included US expectancy ratings for each trial. On the other hand, in the experiment explicitly investigating the role of contingency instructions in classical conditioning (Mertens et al., 2021) described above, the contingency awareness rate was only 53% in a group of participants who received no contingency instructions (compared to 71% in the group with precise, and 91% in the group with general instructions). The ratios observed in the two current experiments are therefore not out of the ordinary.

Self-report measures

Notably, there were no differences in the STAI and BES-A questionnaire results between participants who learned and did not learn the stimulus contingency. Therefore, levels of empathy or anxiety cannot explain the difference in the contingency learning outcome. As a side note, the observers' state anxiety decreased between the beginning and the end of the experiment. The difference was likely caused by elevated anxiety at the beginning of the experiment; when completing the questionnaire at the end of the study, the participants knew they would not experience aversive stimuli.

Even more important, the ratings of the demonstrator and the observational US, referring to the observational learning stage, did not differ between groups. All observers rated the demonstrator's reactions to aversive stimulation as highly natural, expressive

[1]Authors of this study used a differential fear conditioning paradigm with two CS, and gave three groups of participants different instructions. The first group received precise instructions, and was informed which CS would be paired with aversive stimulation and which will not. The second group received general instructions: they were informed that one of the CS would be paired with aversive stimulation, but it would be their task to learn which. The third group received no instructions: they were only informed that they will see two different shapes on the screen and may receive aversive stimulation (just like observers in the current study). Contingency knowledge ratio was lowest in the third group.

and indicative of high discomfort. They also declared that they could identify themselves with the demonstrator and empathised with them. The inclusion of the subjective evaluation of the demonstrator and the observational US has been recommended for observational fear conditioning paradigms which use recordings of more than one demonstrator (Haaker, Golkar et al., 2017). It was especially relevant in the current experiment, in which every participant watched another demonstrator, and the observation happened in real-time. The real-time observation meant that there was no possibility to control the demonstrator's behaviour in the same way as with recordings made in advance.

Finally, high ratings of expressiveness are reassuring from the perspective of evaluating the live observation paradigm. Observational stimuli may arguably be considered weaker than the first-hand experience of electrical stimulation. Nevertheless, even in non-social fear conditioning, researchers have expressed concern related to the intensity or relevance of the stimuli: "Because of ethical restrictions the UCSs, although aversive, are so lacking in intensity or impact that for many individuals these repeated CS–UCS acquisition trials can result in decrease of general arousal due to boredom" (Boucsein et al., 2012).

Psychophysiological responses

In the electrodermal activity analysis, the observational learning stage was characterised primarily by a strong response to the observational US, which did not differ between contingency-aware and contingency-unaware groups (in agreement with the observational US ratings). During observation, the CS reactions were comparatively much lower, and there was no CS+ vs CS– difference in either group. In the subsequent direct-expression stage, only the contingency knowing participants reacted stronger to CS+ than to CS–, while the other group showed no significant difference in CS reactions. Moreover, the between-group differences were manifested in the reactions to CS+. Thus, both groups reacted to the observational US, but only the contingency knowing participants exhibited a typical (Haaker, Golkar et al., 2017) fear-conditioned response.

In the fear-potentiated startle analysis, the observational learning stage was characterised by higher startle responses during both CS than inter-trial intervals, with no group differences. In the direct-expression stage, only the contingency-knowing participants exhibited a fear-conditioned response during the CS+ relative to ITIs. Finding a difference between CS+ and CS– (rather than CS+ and ITI) would have been even more typical for fear conditioning; however, it was not observed, potentially indicating fear generalisation to both CS (Arnaudova et al., 2017). Additionally, there were two trend-level effects: the contingency knowing participants had higher responses during CS– relative to ITI, and lower responses during ITI than the not-knowing participants. This strengthens the conclusion that fear-conditioned responses were present only in the contingency knowing participants. Qualitatively, only this group had a gradation of responses (CS+ > CS– > ITI).

In sum, the results of EDA and FPS analyses are generally aligned in that the conditioning effects were found for contingency knowing participants and were driven mainly by elevated CS+ responses. Interestingly, the results were not the same: the CS+ > CS–

difference was found in EDA, but in FPS, only the CS+ > ITI comparison was significant. However, while both measures are used to evaluate fear conditioning, they rely on different psychophysiological mechanisms (Lonsdorf et al., 2017; Ojala & Bach, 2020).

Relation between contingency knowledge and physiological responses

The pattern of results described above leads to a conclusion that although the observers did respond to the observational US, not all of them formed the CS–US association (reflected in declarative contingency knowledge, skin conductance and startle responses during direct-expression). Is this surprising? The interaction of contingency knowledge with physiological measures has been disputed in the literature. From a theoretical perspective, some authors considered conditioning to be a largely automatic process (LeDoux & Pine, 2016), underlain by a two-system model in which conscious awareness of contingency and conditioned responses arise separately through distinct neural mechanisms. Conversely, other authors argued that awareness is necessary for conditioning to occur (Lovibond & Shanks, 2002), favouring a single process model in which contingency awareness and conditioned responses are closely related. From a physiological perspective, some reports indicated that contingency knowledge is required for conditioned skin conductance responses, but not fear-potentiated startle, e.g. (Hamm & Vaitl, 1996; Sevenster et al., 2014), while others found it necessary for both measures, e.g. (Dawson et al., 2007; Purkis & Lipp, 2001). A recent review (Mertens & Engelhard, 2020) found no evidence of conditioning without awareness and a likely publication bias regarding this subject. The pattern of results observed in the current experiment favours the necessity of contingency knowledge for observational fear conditioning.

Study limitations

The analysis dividing participants based on contingency knowledge was not planned in this study; however, it was inevitable due to the large proportion of contingency unaware participants in the sample. For this reason, the study may have been underpowered for between-group comparisons. The low number of contingency knowing participants can be partially attributed to creating a weak learning situation (vague instructions, low reinforcement ratio). However, weak learning situations may benefit research. On the one hand, they are more likely to reveal meaningful individual differences, and on the other hand, they are more similar to daily life, where ambiguous situations are common (Beckers et al., 2013; Lissek et al., 2006).

The demonstrator's behaviour when aversive stimulation occurred was not entirely spontaneous, as they had received instructions on how to react in an expressive manner. This can be seen as a limitation to the ecological validity, which was necessary to ensure that the aversive stimulation was communicated clearly (this became apparent during pilot experiments). However, within the bounds set by the instructions, each demonstrator behaved differently, and some were more expressive than others. Crucially, median ratings of both naturalness and expressiveness of demonstrator's reactions were high. The ratings confirm that it is feasible to use the real-time procedure. Another lim-

itation to the naturalness of the situation is that the participants were seated in separate rooms, and a video stream was used instead of direct observation. However, given that the participants started the experiment together, the aspect of engagement was maintained. A video streaming solution is also easier to implement, for example, in the fMRI environment[2].

4.2 Conclusions from the neuroimaging experiment

The goal of the neuroimaging experiment was to test whether friendship enhances observational fear conditioning. Physiological responses and brain activation patterns were compared directly between participants who observed their friends and participants who observed persons they did not know. First, questionnaires and physiological measures will be discussed. Concerning the fMRI results, it is reasonable to evaluate the main effects of the task before proceeding to between-group comparisons, especially in a novel procedure. Consequently, the discussion of fMRI results will focus first on describing the shared patterns of brain activity, revealed by analysing both groups of participants together, and then group comparisons will be discussed.

Declarative measures

Three-quarters of participants correctly identified the contingency of CS+ and US in the questionnaire at the end of the experiment. This rate of declarative learning was higher than in the psychophysiological experiment. Both experiments used the same instructions, reinforcement ratio, and contingency assessment procedure. However, the neuroimaging experiment did not include startle presentation and was conducted in a different laboratory, where the demonstrator's surroundings were arranged to be distractor-free. These two factors may have contributed to the higher contingency knowledge ratio in the neuroimaging experiment than the psychophysiological one. Crucially though, the ratio of contingency knowing participants did not differ significantly between friend and stranger groups. Therefore, on a declarative level learning from friends was equally effective as from strangers.

Observers from the friend and stranger groups had similar traits of empathy and anxiety: no significant differences were found in the results of the STAI and SWE questionnaires. Likewise, there were no significant effects of the group in the state anxiety questionnaires. Compared to the psychophysiological experiment, there was, interestingly, no change in state anxiety between the beginning and the end of the experiment (previously, a slight decrease was observed). Additionally, the neuroimaging experiment used a different but arguably more accurate questionnaire to measure empathy (SWE instead of BES-A).

Observers assessed the behaviour of the demonstrator receiving aversive stimulation at the end of the experiments. These ratings are essential for evaluating the procedure's

[2] That being said, with enough creativity the demonstrator could be placed in the magnet room — but it would be technically challenging.

effectiveness and comparing participants' perceptions of the observational US between groups. The majority of participants rated the perceived discomfort, naturalness, and expressiveness highly. Participants also declared that they empathised with the demonstrator to a large degree. These ratings show that the demonstrator's reactions successfully conveyed information about threat. Crucially, there was no difference in any of the ratings between friend and stranger groups, which suggests that the threat information was received similarly on a declarative level.

Skin conductance responses

The skin conductance responses link the neuroimaging and psychophysiological experiments, even though the methods used for acquisition and analysis were different. In the observational learning stage, the observational US responses were significantly higher than the US omission (no US), confirming that the observers reacted strongly to the threat stimulus. Similarly, in the direct-expression, the responses to CS+ were significantly higher than to the CS–, indicating a fear-conditioned response. The group effect and the group × stimulus interaction were not significant, suggesting a lack of between-group differences in physiological responses. Interestingly, Bayesian ANOVA indicated overwhelming evidence only for the main effect of stimulus in observational learning and inconclusive evidence regarding the other effects and interactions. The discrepancy is probably caused by a large variability of skin conductance responses between participants. Consequently, the physiological results should be interpreted with caution.

Observational US – activation analysis

The observational learning stage activation analysis was centred around the observational US because this stimulus (another person reacting to the aversive stimulation paired with CS+) is the most important cue for threat learning (Haaker, Golkar et al., 2017; Lindström et al., 2018). The US > noUS contrast was first evaluated for all subjects. The results show an extensive activation of several brain regions, which can be grouped into two categories: related to fear and related to social perception.

The core network related to fear contains the amygdala, anterior part of the insular cortex and the anterior cingulate cortex. All three regions were activated in the observational learning task. The same regions have been characterised as a cross-modal (self/other) learning network in an fMRI study comparing direct and social fear conditioning in a within-subject design (Lindström et al., 2018). Results from the current experiment are consistent with this description. The classical view of the amygdala role in fear conditioning is that it is a site of convergence of CS and US information, and that it participates in associative memory formation and reactivation (LeDoux, 2003; Maren & Fanselow, 1996; Phelps & LeDoux, 2005). Moreover, it is also involved in recognising fearful facial expressions (Adolphs et al., 1995). While the amygdala is frequently highlighted, the AI and ACC play an equally important role and are consistently activated in fear conditioning experiments (Fullana et al., 2016). The two structures are associated with autonomic and interoceptive processes (Craig, 2009; Critchley & Harrison, 2013;

Medford & Critchley, 2010), and in fear conditioning, they are likely responsible for representation of the aversive value of pain, both self-experienced and empathetic (Lamm et al., 2011; Lindström et al., 2018).

Another group of activated regions contains structures associated with perception of dynamic social stimuli: supplementary motor cortex, fusiform gyrus (including the FFA) and STS. The demonstrator's reaction to the aversive stimulation (observational US) involved hand twitching, facial grimace, and, sometimes, upper body movement. Consequently, the observed activation of the supplementary motor cortex likely reflects movement mirroring (Keysers & Gazzola, 2009). The STS is involved in analysing social stimuli signalling actions of other individuals and is activated by both biological motion and static images (Allison et al., 2000). Moreover, its posterior part is functionally connected to regions crucial for social perception (e.g. fusiform gyrus, orbitofrontal cortex), action observation (e.g. inferior parietal lobule, inferior frontal gyrus) and theory of mind (e.g. temporoparietal junction, medial prefrontal cortex, posterior cingulate cortex, precuneus). It has been proposed as a hub that integrates these low- and high-level processes (Yang et al., 2015). Finally, the FFA is an area specialised in the perception of faces (Kanwisher et al., 1997). That being said, the perception of faces and emotional expression recognition is not constrained to FFA and involves a network that also includes the pSTS and the amygdala.

In summary, observational learning was related to activity in regions typically associated with fear and social perception. The two sets are not exclusive: the amygdala, for example, is associated with both fear conditioning and facial expression recognition. The pattern of results obtained during acquisition is consistent with a recent review of social fear learning (Olsson et al., 2020). It confirms that observational and self-experienced fear conditioning partially share their neural underpinnings. Crucially, these effects were observed for both groups of participants and there were no differences between friends and strangers.

Direct CS – activation analysis

The activation analysis of the direct-expression stage focused on the reactions to CS presentations. The fear-conditioned response was tested by evaluating the CS+ > CS– contrast, first for all subjects together. The results show an activation of the AI and aMCC, two regions that were also activated in the US > noUS contrast during observation. This result conforms with expectations (Fullana et al., 2016; Olsson et al., 2007).

Interestingly, though, the amygdala was not activated in this contrast (neither with whole-brain nor with small volume corrected thresholds). This can be surprising because, as discussed above, multiple studies have demonstrated the amygdala's involvement both in fear acquisition and expression. Moreover, an early fMRI study of observational fear conditioning (Olsson et al., 2007) demonstrated that the amygdala was activated, in addition to the AI and the ACC, during both observational learning and direct-expression. Conversely to the current results, this study also reported that the amygdala activation had a similar extent during both stages, while AI and ACC activa-

tions were more extensive during direct-expression. However, the mentioned study used a smaller sample size and lenient statistical thresholds than the current experiment.

The lack of amygdala activation might have been attributed to quick habituation. Indeed, studies often report amygdala activations primarily during early stages of the task (Fullana et al., 2016). Moreover, the habituation effect may have been enhanced by the fact that the direct-expression stage contained no US presentations; in this sense, it was an extinction stage as much as a test stage. To evaluate the effects of extinction, the temporal modulation of CS+ responses was tested. However, no modulation effects were found in the amygdala, thus the result does not support the habituation hypothesis.

It is worth noting that despite evidence linking the amygdala with fear, recent neuroimaging meta-analyses of fear conditioning (Fullana et al., 2016) and extinction (Fullana et al., 2018) studies did not find robust evidence for amygdala activation. Authors of the meta-analyses propose that the result most likely reflects a limitation of fMRI related to its temporal and spatial resolution (and the necessity of using multiple trial repetitions). In rodents, the CS–US association is encoded by a small number of sparsely distributed neurons which inhibit their neighbours (Krabbe et al., 2018; Reijmers et al., 2007). It has been proposed that in humans, responses generated from such a sparse organisation would not be detectable using a mass-univariate activation approach (as used in the current study) but should still allow response classification using multivariate methods (Bach et al., 2011). Alternatively, it could be speculated that in the current study, the amygdala was only involved during fear acquisition, while fear expression was delegated to other structures (AI, aMCC). A third interpretation, also proposed by Fullana et al. (2018), would assume that amygdala activation is typical only for intense fear states, and although anticipatory CS+ responses were observed during direct-expression (in skin conductance and neuroimaging results), they did not necessarily indicate strong fear.

Returning to the temporal modulation effects, although they were not observed in the amygdala, the analysis for both groups together found a cluster in the aMCC where CS+ responses decreased with time. This finding complements the activation of aMCC described previously and further highlights the aMCC involvement during direct-expression.

The temporal modulation effects were also compared between friend and stranger groups, and surprisingly, this was the only between-group comparison that yielded significant results. However, the cluster locations (which do not correspond to activations found in previous analyses) and the ambiguous nature of the evaluated contrast (friend CS+ × time > stranger CS+ × time) make this result difficult to interpret. Therefore, it can be concluded that there were no major differences between the groups concerning the temporal dynamics of fear extinction.

Psychophysiological interactions

The final stage of the fMRI analysis used a psychophysiological interaction approach. Conceptually, such analysis focuses on identifying regions that change their connectivity in a task-dependent manner. The analysis was conducted for both stages of the experiment. In the observational learning stage, the US > noUS contrast was used to define

the psychological variable. The analysis with both groups together used AI and rpSTS seeds. The between-group comparisons used AI, rpSTS, rTPJ, FFA, aMCC, and amygdala seeds. In direct-expression, the CS+ > CS– contrast was used to define the psychological variable; AI and aMCC (the two regions activated in this task) were used as seeds for all comparisons. However, significant results were observed only for observational learning and only in the analysis with both groups together.

The PPI analysis was used in one previous study of observational fear conditioning, in which subjects completed the conditioning task through observation and first-hand experience (Lindström et al., 2018). That study demonstrated that during social fear learning, the AI increased its connectivity with the TPJ following the US presentations. Consequently, the AI seed region and US > noUS contrast were chosen as the primary comparison in the current study. Both groups of participants were first analysed together to identify effects shared by friends and strangers. Contrary to expectations, significant effects were not observed in the TPJ. Instead, increased connectivity with the AI was found in the rpSTS, a region that, like TPJ, is involved in social processing (Yang et al., 2015). Because the cluster which showed significant effects in this analysis extended largely beyond the rpSTS, an analogous analysis was conducted with the rpSTS seed defined independently of the previous result. The analysis showed enhanced connectivity of rpSTS not just with the AI but also with the fusiform gyrus and the amygdala.

All four regions were also activated in the US > noUS contrast. However, the PPI analysis is designed to detect interaction effects beyond the coactivation (Di et al., 2020; O'Reilly et al., 2012). Therefore, the PPI results complement the activation results and show that a region characteristic for perception and processing of social stimuli (rpSTS) interacts not only with other socially relevant areas (FFA) but also those engaged by fear (amygdala, AI).

Finally, the PPI effects were compared between friend and stranger groups to address the possibility that the group differences could be expressed in the connectivity between the key regions rather than their overall recruitment. However, no significant between-group differences were found for the AI and rpSTS seeds. To test the issue more extensively, analogous analyses were performed for the remaining regions of interest, namely the amygdala, aMCC, rTPJ and FFA, again without obtaining significant group differences. In conclusion, it does not seem likely that the groups differ in task-related connectivity. However, it has to be kept in mind that the PPI analysis is, by its nature, less sensitive than the activation analysis (O'Reilly et al., 2012) because it tests subtler effects.

Study limitations

The findings of this study need to be seen in the light of some limitations. First, the experimental paradigm closely followed the one used in the psychophysiological experiment. Consequently, the same considerations apply here: using video transmission and instructing the demonstrators on how to react to aversive stimulation was a necessary compromise between a fully spontaneous interaction and a controlled experiment. However, these factors do not appear to be severely limiting: as previously, each demon-

strator behaved differently (within the bounds set by the instructions), and the observers rated the naturalness and expressiveness of the demonstrator's reactions highly.

Second, although the friend and stranger groups were closely matched in terms of the procedure and the stimuli (the stranger group watched recordings from the friend group), the circumstances surrounding the experiment differed. In the friend group, participants started the experiment together, and in many cases, they also travelled to the laboratory together. This may have weakened their aversive reactions due to social buffering. However, it has to be noted that, for example, the state anxiety scores did not differ between groups (social buffering will be discussed more thoroughly in the general discussion). It can also be argued that because they arrived at the laboratory together, the situation was more socially engaging for friends, making the groups less comparable. At the same time, this would only enhance the friendship effect, which was not detected in the study.

Finally, the study included only male subjects, which made the sample more homogenous but potentially limited the generalizability of the findings. Because the psychophysiological experiment also included only male participants, potential sex differences will be further considered in the general discussion.

4.3 Common themes

The psychophysiological and neuroimaging experiments discussed above investigated the mechanisms of fear conditioning through observation of friends or strangers. The first experiment validated the experimental paradigm based on live observation in friend dyads and highlighted the importance of declarative contingency knowledge. The second experiment extended the first and compared friend and stranger groups in an fMRI paradigm. So far, these results have been discussed mainly in the context of emotional contagion and compared to previous studies of fear conditioning, both through observation and first-hand experiences.

This section will highlight themes that are common for both experiments and cannot be ignored when interpreting the results. I will start by summarising the advantages of the live observation approach developed in this research project. Next, I will compare the current research to studies on empathy for pain, which is an important line of research related to the perception of other people's negative emotions, separate from fear conditioning. Then, I will propose that to explain the lack of between-group differences, the friend-stranger comparison should be considered not only through the lens of empathy but also in terms of the value of the social information being transferred. In the end, I will discuss the phenomenon of social buffering, which is relevant when pairs of people participate in an experiment together. I will also speculate on the possibility of generalising the obtained results from male friendship to friendship in general.

Developing a naturalistic paradigm

The design of the experiments described in the thesis was based on the observational fear conditioning paradigm (Haaker, Golkar et al., 2017). However, significant changes have

been made to improve its ecological validity. Crucially, a decision was made that pairs of participants should take part in the experiment together. Establishing the modified paradigm presented a set of challenges, both conceptual (e.g. which parts of the procedure should be modified), technical (e.g. how to solve the problem of video streaming and stimulus presentation) and practical (e.g. what instructions should the demonstrators receive). Due to the scope of the changes, it was essential for this research project that the psychophysiological experiment was conducted before the neuroimaging experiment, with a smaller group of participants. Consequently, the neuroimaging experiment design could utilise the experience gained previously.

The procedure developed in the course of the current research project has unique advantages. First, the experiment carried out with both participants has the potential to be more engaging, in line with the desire to improve the ecological validity of social studies (Matusz et al., 2019; Neisser, 1980). Second, in this setting, it becomes easier to design experiments around existing relationships between the participants. The current studies investigated the social transfer of fear in pairs of friends. Future research could investigate other types of dyads, for example, comparing between familiarity and familiality (friends vs family members) or replicating the comparisons of social ingroups and outgroups (Golkar & Olsson, 2017). Finally, following the same principles, the procedure can be applied with different research methods: psychophysiology or neuroimaging. Perhaps, even, there can be some cross-talk with other areas of study, such as empathy for pain.

Comparison to studies on empathy for pain

In the observational fear learning stage, the observers watched the demonstrators who exhibited reactions of discomfort after an aversive electrical stimulation, which was paired with a visual stimulus. The stimulation was not painful, but the observers did not have first-hand experience with it and could only interpret the demonstrator's reaction. Thus, a comparison with studies on empathy for pain seems justified.

It is widely accepted that empathy for pain engages regions of the brain that are also responsible for first-hand painful experiences, of which AI and ACC are the most important and are consistently reported (Bernhardt & Singer, 2012; Lamm et al., 2011; Zaki et al., 2016). Additional evidence for empathy being grounded in the same neural mechanisms (including neurotransmitter activity) as first-hand pain was provided by studies showing that inducing pain analgesia also reduces pain empathy (Rütgen et al., 2015). Significant activations to witnessing pain are not limited to areas related to processing negative emotions but also occur in somatosensory regions (Keysers et al., 2010; Riečanský & Lamm, 2019). Therefore, there is considerable overlap between patterns of activation reported for empathic pain and those reported for first-hand (Fullana et al., 2016) and observational (Olsson et al., 2020) conditioning.

The distinction between strangers and close others has also been explored in the context of empathy for pain. For instance, in one study (Cheng et al., 2010), participants watched short sequences of pictures depicting pain inflicted on hands or feet (such as hitting one's toe). The pain stimuli were preceded with priming photo cues from three

conditions: self, loved one, or stranger. Participants were asked to imagine the situation happening to the person in the picture. The strength of activations seen in the ACC, insula, and rTPJ was similar in the self and loved one conditions and different in the stranger condition. In the direct comparisons, the ACC was significantly more active for loved ones than for strangers, and the rTPJ was more active for strangers than loved ones[3]. Another study (Wang et al., 2016) used a similar paradigm, with priming pictures of close friends and strangers, in an EEG setting. Its results showed differences between friend and stranger conditions in two components of evoked potentials, both early (N110, localised to the anterior prefrontal cortex; related to automatic processing) and late (posterior P300 and late positive components; related to cognitive appraisals). Based on these results, it could be expected that reactions to the observational US in the current experiment should be higher in the friends group, which was not the case.

There are, however, two crucial distinctions between the current experiments and the studies on empathy for pain. The first is methodological: most studies on the empathy for pain use stimuli with different content and temporal dynamics. Although some of the discussed studies did use short reactions to electrical stimulation as pain stimuli (Rütgen et al., 2015), others presented pictures of limbs getting hurt (Cheng et al., 2010; Wang et al., 2016) or videos showing reactions of actors to accidents from everyday life (Jankowiak-Siuda et al., 2019; Jankowiak-Siuda et al., 2015). In the last example, the actors' reactions were purposefully drawn out and lasted three seconds. The second and arguably more significant difference is the context: in the discussed studies on pain, participants were asked to watch the situation or imagine it happening. In the current experiments, the participants did not only watch the situation but were also led to believe that the same situation would happen to them in the following part of the experiment. Therefore, the aspect of increased self-relevance was equally present in the friend and stranger condition. Moreover, since the observers received vague instructions about the task specifics and did not have prior experience with the electrical stimulation, their focus was probably on identifying the threat. In summary, the observer's responses to the observational US certainly carry a component of empathy for pain. However, they should be considered primarily in the context of fear conditioning and social learning of threats.

Familiarity, similarity and informational value

It has been proposed that empathy affects emotional learning, and emotional learning affects empathy (Olsson & Spring, 2018). The Russian doll model of empathy and its underlying perception-action model suggest that familiarity should modulate emotional contagion and, more broadly, empathy. Greater familiarity or similarity with the

[3]It has to be noted that the statistical threshold used by this study can be seen as lenient. Table S2 in the described article lists the following statistics of peak activations: $t(35) = 2.69$ in ACC for loved one > stranger, $t(35) = 2.92$ in rTPJ for stranger > loved one. This corresponds to uncorrected p-values of 0.005 and 0.003, which would not be sufficient to pass corrections for multiple comparisons used in this thesis (e.g. $p = 0.001$ was used as a cluster defining threshold for cluster size correction, in line with current recommendations). That being said, authors additionally support their claims with a region of interest analysis.

demonstrator should mean that a richer representation (formed through multiple interactions) will be activated when witnessing their emotional state (Preston & de Waal, 2002). In line with these assumptions, studies discussed previously in this thesis showed greater empathy for the pain of close ones (Cheng et al., 2010) and modulation of fear conditioning by racial and social group membership (Golkar et al., 2015; Golkar & Olsson, 2017), reflected in neuroimaging or psychophysiological measures. However, seemingly at odds with the literature, the observational fear conditioning experiment conducted as part of this thesis showed no differences between the groups of friends and strangers.

More specifically, the groups were compared in several aspects. No differences were found in self-report measures of contingency knowledge or the perception of the demonstrator and the observational US. No effects of the group were found in the skin conductance responses. Finally, no differences were found in the fMRI analysis of observational US and direct CS responses in activation and connectivity frameworks. Moreover, Bayesian analysis of region-of-interest activations mostly indicated moderate evidence of absence of effects.

However, the current study did not attempt to emphasise the friend vs stranger distinction. For comparison, the study on racial and social group biases (Golkar & Olsson, 2017) recruited volunteers who were ardent supporters of competing local football clubs and explicitly instructed them about the affiliation of the demonstrator. The current study relied on implicit differences instead: observers watched either the person they knew very well or the person they did not know at all. Other than that, even in the stranger group, the observer and demonstrator were similar: they were white men of a similar age, wore similar clothing, and were likely to have a similar socioeconomic status (most participants were students living in Warsaw). As a result, the differences in the current design were reduced to familiarity itself.

At this point, it may be worthwhile to take a step back from explanations based on empathy and consider the friend - stranger comparison from a broader perspective of social learning strategies. Laland (2004) proposed that the strategies employed by social animals, including humans, when they learn from each other can be divided into two categories. The first category contains heuristics for when social learning is preferred, such as "copy when asocial learning is costly" or "copy when uncertain". The second category contains heuristics for who is the preferred source of information, such as "copy kin", "copy friends", or "copy older individuals". A question can then be asked, to what degree are these strategies applicable to the experimental situation created for each group of participants (or, more precisely, which factors influenced the value of the threat information, as the participants could not choose whom they learned from). I would like to argue that the observers in the stranger group had little reason to value the information less than the friends group.

Learning about threats encourages social learning, as direct experiences may be costly. When participants were told that they would undergo the same procedure as the demonstrator, they were probably inclined to learn what it would mean for them. The context (corresponding to "when" strategies) may have been more important as the source of the information. Regarding the "who" strategies, trusting strangers does not have to be

much less viable than trusting friends, especially in the context of threat and in the absence of additional clues about the stranger (i.e. as long as they can be seen as in-group individuals).

Social buffering

The experiments were conducted with pairs of friends, who were present at the laboratory together. Although the goal of the procedure was to study emotional contagion, such conditions may have also facilitated an opposite phenomenon – social buffering. Social buffering is a process in which the presence of another person (or conspecific, as this process is evolutionary conserved among social animals) reduces the behavioural and physiological responses to aversive events (Ditzen & Heinrichs, 2014; Kikusui et al., 2006; Oliveira & Faustino, 2017). Apart from humans, social buffering has been demonstrated in several species, from rats (Davitz & Mason, 1955; Fuzzo et al., 2015) to non-human primates (Wittig et al., 2016).

Directly relevant to the current experiments, one recent study showed that the mere presence of another, unknown person, who did not provide active support, caused lower skin conductance responses in participants exposed to aversive sounds (such as human screams) compared to the participants who were in the room alone (Qi et al., 2020). Another study investigated the role of social buffering in threat and safety learning and showed that fear extinction was more efficient when two participants, who sat nearby and could see each other, underwent extinction training simultaneously (Pan et al., 2020).

Is it possible that the current experiments found no differences between friends and strangers because of two opposing mechanisms: response enhancement (due to familiarity with the demonstrator) and attenuation (due to social buffering)? The answer to that question can only be based on speculation. Admittedly, the influence of social buffering cannot be entirely excluded; however, it is unlikely that it was large enough to compensate for the hypothetical effects of familiarity. First, the state anxiety questionnaire results do not indicate any difference between friend and stranger groups. Second, unlike in the two studies mentioned in the previous paragraph, the demonstrator and observer were not in the same room. The observer could see the demonstrator during the observational learning stage but not during the direct-expression. What is more, the procedure was identical (except the replacement of the video stream by a recording) in the stranger group, so the buffering effects would need to carry over from before the experiment to cause a difference. Finally, during observation, the friend was the subject of a threatening situation rather than a passive bystander or an active supporter, which probably favours emotional contagion rather than buffering. In conclusion, social buffering is an interesting phenomenon, also in the context of observational fear conditioning. However, it is unlikely to have played a major role in shaping the results in the current paradigm.

Sex differences and gender relations

The current experiments were conducted with heterosexual male participants only. Considering the novelty of the research topic, such a decision was made to eliminate an additional source of inter-subject variability. This leaves open the question of whether the same results would be observed for women or a mixed sample of participants. Of course, the question can only be answered by future research. However, existing literature can be used as a basis for speculation.

First, are there essential sex differences in emotional responses, particularly in the context of fear? Reviews indicate that men and women respond differently to female and male emotional expressions and show different empathy for pain. Broadly speaking, although women are better at recognising and expressing emotions, men show greater responses to threatening cues; such response patterns may have been shaped both culturally and evolutionally (Kret & De Gelder, 2012; Proverbio, 2021). However, the authors of the reviev emphasise that care should be taken to avoid interpreting results only in a stereotype-consistent way (that women are more emotional than men), suggesting that while emotional stimuli signalling threat might be more distressing for women, for men they may be more behaviourally relevant and elicit a more potent orienting response (Kret & De Gelder, 2012). The effects may depend not only on who is the observer but also on who is being observed. For example, one study investigated empathic responses of male and female participants to videos of actors experiencing pain (Jankowiak-Siuda et al., 2015). An ROI analysis reported in this study has found an interaction effect between actor's sex and attractiveness in the activations of several regions of interest, including the AI and ACC. Participants showed greater responses to the pain of an attractive woman (compared to a less attractive woman) and greater responses to a less attractive man (compared to an attractive man). A similar interaction effect has been found in a similar study using facial EMG to measure muscle activity related to empathy for pain (Jankowiak-Siuda et al., 2019).

Second, do male friendships differ from other kinds of friendship? This is a question about gender relations rather than sex differences. Male friendships are generally characterised by less overt affection than female friendships (Demir & Orthel, 2011), and the affection can often be shown through playful derogation rather than directly (McDiarmid et al., 2017). That being said, one needs to avoid stereotypical thinking (that friendships among men are of low quality) because male friendships are emotional, affectionate and, indeed, intimate (Kaplan & Rosenmann, 2014; Levy, 2005; Robinson et al., 2019).

Finally, what about previous studies of observational fear conditioning? While some neuroimaging studies also investigated a male-only sample (Haaker, Yi et al., 2017), there were also neuroimaging (Lindström et al., 2018) and psychophysiological (Olsson et al., 2016; Pärnamets et al., 2020; Williams & Conway, 2020) experiments which included male and female participants together. These studies matched the observer and demonstrator by gender, but crucially none reported differences in responses between males and females.

To summarise, males process emotional information differently than females, also

on a physiological level, and male friendships are subject to specific masculine rules. Therefore, some caution is advisable when generalising the results. At the same time, male friendships do not exclude emotionality, and previous studies on observational fear conditioning suggest that the principal neuronal mechanisms of observational fear conditioning would likely be the same if a female or a mixed group of participants were investigated.

Concluding remarks

Results obtained in the experiments presented in the thesis lead to the conclusion that learning about threats from friends and strangers is equally effective. That being said, the influence of social factors on observational fear conditioning in particular, and social learning in general, is a fascinating subject. Social relations are complex, and friendship between men is just one element of the social landscape. Certainly, further research will shed more light on the various influences on social learning about threats.

Bibliography

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B. & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.*, *8*.

Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. R. (1995). Fear and the human amygdala. *J. Neurosci.*, *15*(9), 5879–5891.

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annu. Rev. Psychol.*, *60*, 693–716.

Adriaense, J. E. C., Koski, S. E., Huber, L. & Lamm, C. (2020). Challenges in the comparative study of empathy and related phenomena in animals. *Neurosci. Biobehav. Rev.*, *112*, 62–82.

Alcalá-López, D., Smallwood, J., Jefferies, E., Van Overwalle, F., Vogeley, K., Mars, R. B., Turetsky, B. I., Laird, A. R., Fox, P. T., Eickhoff, S. B. & Bzdok, D. (2018). Computing the social brain connectome across systems and states. *Cereb. Cortex*, *28*(7), 2207–2232.

Allison, T., Puce, A. & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends Cogn. Sci.*, *4*(7), 267–278.

American Psychological Association. (n.d.). *False-belief task*. Retrieved July 21, 2021, from https://dictionary.apa.org/false-belief-task

Andraka, K., Kondrakiewicz, K., Rojek-Sito, K., Ziegart-Sadowska, K., Meyza, K., Nikolaev, T., Hamed, A., Kursa, M., Wójcik, M., Danielewski, K., Wiatrowska, M., Kublik, E., Bekisz, M., Lebitko, T., Duque, D., Jaworski, T., Madej, H., Konopka, W., Boguszewski, P. M. & Knapska, E. (2021). Distinct circuits in rat central amygdala for defensive behaviors evoked by socially signaled imminent versus remote danger. *Curr. Biol.*, *31*(11), 2347–2358.e6.

Arnaudova, I., Krypotos, A.-M., Effting, M., Kindt, M. & Beckers, T. (2017). Fearing shades of grey: Individual differences in fear responding towards generalisation stimuli. *Cogn. Emot.*, *31*(6), 1181–1196.

Askew, C. & Field, A. P. (2007). Vicarious learning and the development of fears in childhood. *Behav. Res. Ther.*, *45*(11), 2616–2627.

Atsak, P., Orre, M., Bakker, P., Cerliani, L., Roozendaal, B., Gazzola, V., Moita, M. & Keysers, C. (2011). Experience modulates vicarious freezing in rats: A model for empathy. *PLoS One*, *6*(7), e21855.

Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.*, *12*(1), 26–41.

Bach, D. R., Daunizeau, J., Friston, K. J. & Dolan, R. J. (2010). Dynamic causal modelling of anticipatory skin conductance responses. *Biol. Psychol.*, *85*(1), 163–170.

Bach, D. R., Flandin, G., Friston, K. J. & Dolan, R. J. (2009). Time-series analysis for rapid event-related skin conductance responses. *J. Neurosci. Methods*, *184*(2), 224–234.

Bach, D. R., Flandin, G., Friston, K. J. & Dolan, R. J. (2010). Modelling event-related skin conductance responses. *Int. J. Psychophysiol.*, *75*(3), 349–356.

Bach, D. R., Weiskopf, N. & Dolan, R. J. (2011). A stable sparse fear memory trace in human amygdala. *J. Neurosci.*, *31*(25), 9383–9389.

Bandura, A. & Rosenthal, T. L. (1966). Vicarious classical conditioning as a function of arousal level. *J. Pers. Soc. Psychol.*, *3*(1), 54–62.

Beckers, T., Krypotos, A.-M., Boddez, Y., Effting, M. & Kindt, M. (2013). What's wrong with fear conditioning? *Biol. Psychol.*, *92*(1), 90–96.

Ben-Ami Bartal, I., Decety, J. & Mason, P. (2011). Empathy and pro-social behavior in rats. *Science*, *334*(6061), 1427–1430.

Berger, S. M. (1962). Conditioning through vicarious instigation. *Psychol. Rev.*, *69*(5), 450–466.

Bernhardt, B. C. & Singer, T. (2012). The neural basis of empathy. *Annu. Rev. Neurosci.*, *35*, 1–23.

Biswal, B. B. (2012). Resting state fMRI: A personal history. *Neuroimage*, *62*(2), 938–944.

Biswal, B. B., Yetkin, F. Z., Haughton, V. M. & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.*, *34*(4), 537–541.

Blechert, J., Michael, T., Williams, S. L., Purkis, H. M. & Wilhelm, F. H. (2008). When two paradigms meet: Does evaluative learning extinguish in differential fear conditioning? *Learn. Motiv.*, *39*(1), 58–70.

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V. & van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*(1), 1–15.

Boccadoro, S., Cracco, E., Hudson, A. R., Bardi, L., Nijhof, A. D., Wiersema, J. R., Brass, M. & Mueller, S. C. (2019). Defining the neural correlates of spontaneous theory of mind (ToM): An fMRI multi-study investigation. *Neuroimage*, *203*, 116193.

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D. & Beckers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. *J. Behav. Ther. Exp. Psychiatry*, *44*(2), 201–206.

Bottenhorn, K. L., Flannery, J. S., Boeving, E. R., Riedel, M. C., Eickhoff, S. B., Sutherland, M. T. & Laird, A. R. (2019). Cooperating yet distinct brain networks engaged during naturalistic paradigms: A meta-analysis of functional MRI results. *Netw Neurosci*, *3*(1), 27–48.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., Filion, D. L. & Society for Psychophysiological Research Ad Hoc Committee

on Electrodermal Measures. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034.

Brett, M. (2016). *Thresholding with random field theory*. https://matthew-brett.github.io/teaching/random_fields.html

Cambiaghi, M. & Sacchetti, B. (2015). Ivan petrovich pavlov (1849-1936). *J. Neurol.*, *262*(6), 1599–1600.

Carré, A., Stefaniak, N., D'Ambrosio, F., Bensalah, L. & Besche-Richard, C. (2013). The basic empathy scale in adults (BES-A): Factor structure of a revised form. *Psychol. Assess.*, *25*(3), 679–691.

Casetta, G., Nolfo, A. P. & Palagi, E. (2021). Yawn contagion promotes motor synchrony in wild lions, panthera leo. *Anim. Behav.*, *174*, 149–159.

Cheng, Y., Chen, C., Lin, C.-P., Chou, K.-H. & Decety, J. (2010). Love hurts: An fMRI study. *Neuroimage*, *51*(2), 923–929.

Cox, R. W. & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR Biomed.*, *10*(4-5), 171–178.

Craig, A. D. B. (2009). How do you feel–now? the anterior insula and human awareness. *Nat. Rev. Neurosci.*, *10*(1), 59–70.

Critchley, H. D. & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron*, *77*(4), 624–638.

Cuff, B. M. P., Brown, S. J., Taylor, L. & Howat, D. J. (2016). Empathy: A review of the concept. *Emot. Rev.*, *8*(2), 144–153.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *J. Pers. Soc. Psychol.*, *44*(1), 113–126.

Davitz, J. R. & Mason, D. J. (1955). Socially facilitated reduction of a fear response in rats. *J. Comp. Physiol. Psychol.*, *48*(3), 149–151.

Dawson, M. E., Rissling, A. J., Schell, A. M. & Wilcox, R. (2007). Under what conditions can human affective conditioning occur without contingency awareness? test of the evaluative conditioning paradigm. *Emotion*, *7*(4), 755–766.

Demir, M. & Orthel, H. (2011). Friendship, real-ideal discrepancies, and well-being: Gender differences in college students. *J. Psychol.*, *145*(3), 173–193.

de Waal, F. B. M. (2009). *The age of empathy*. Harmony Books.

de Waal, F. B. M. & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nat. Rev. Neurosci.*, *18*(8), 498–509.

Di, X., Zhang, Z. & Biswal, B. B. (2020). Understanding psychophysiological interaction and its relations to beta series correlation. *Brain Imaging Behav.*

Ditzen, B. & Heinrichs, M. (2014). Psychobiology of social support: The social dimension of stress buffering. *Restor. Neurol. Neurosci.*, *32*(1), 149–162.

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Salo, T., Valabregue, R., Amlien, I. K., Liem, F., … Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright,

J., Durnez, J., Poldrack, R. A. & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nat. Methods*, *16*(1), 111–116.

Fanselow, M. S. & Wassum, K. M. (2015). The origins and organization of vertebrate pavlovian conditioning. *Cold Spring Harb. Perspect. Biol.*, *8*(1), a021717.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R. & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage*, *47, Supplement 1*, S102.

Fox, M. D. & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.*, *8*(9), 700–711.

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A. & Radua, J. (2016). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Mol. Psychiatry*, *21*(4), 500–508.

Fullana, M. A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., Radua, J. & Harrison, B. J. (2018). Fear extinction in the human brain: A meta-analysis of fMRI studies in healthy participants. *Neurosci. Biobehav. Rev.*, *88*, 16–25.

Fuzzo, F., Matsumoto, J., Kiyokawa, Y., Takeuchi, Y., Ono, T. & Nishijo, H. (2015). Social buffering suppresses fear-associated activation of the lateral amygdala in male rats: Behavioral and neurophysiological evidence. *Front. Neurosci.*, *9*, 99.

Gerster, S., Namer, B., Elam, M. & Bach, D. R. (2018). Testing a linear time invariant model for skin conductance responses by intraneural recording and stimulation. *Psychophysiology*, *55*(2).

Golkar, A., Castro, V. & Olsson, A. (2015). Social learning of fear and safety is determined by the demonstrator's racial group. *Biol. Lett.*, *11*(1), 20140817.

Golkar, A., Haaker, J., Selbing, I. & Olsson, A. (2016). Neural signals of vicarious extinction learning. *Soc. Cogn. Affect. Neurosci.*, *11*(10), 1541–1549.

Golkar, A. & Olsson, A. (2017). The interplay of social group biases in social threat learning. *Sci. Rep.*, *7*(1), 7685.

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J. M., Coalson, R. S., Nguyen, A. L., McDermott, K. B., Shimony, J. S., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Nelson, S. M. & Dosenbach, N. U. F. (2017). Precision functional mapping of individual human brains. *Neuron*, *95*(4), 791–807.e7.

Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L. & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.*, *5*, 13.

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Jarecka, D., Notter, M. P., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Goncalves, M., Jordan, K., Waskom, M., Wong, J., Modat, M., Loney, F., … Ghosh, S. (2019). Nipy/nipype: 1.2.0.

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P. & Citi, L. (2016). cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Trans. Biomed. Eng.*, *63*(4), 797–804.

Greve, D. N. & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*(1), 63–72.

Haaker, J., Golkar, A., Selbing, I. & Olsson, A. (2017). Assessment of social transmission of threats in humans using observational fear conditioning. *Nat. Protoc.*, *12*(7), 1378–1386.

Haaker, J., Maren, S., Andreatta, M., Merz, C. J., Richter, J., Richter, S. H., Meir Drexler, S., Lange, M. D., Jüngling, K., Nees, F., Seidenbecher, T., Fullana, M. A., Wotjak, C. T. & Lonsdorf, T. B. (2019). Making translation work: Harmonizing cross-species methodology in the behavioural neuroscience of pavlovian fear conditioning. *Neurosci. Biobehav. Rev.*, *107*, 329–345.

Haaker, J., Yi, J., Petrovic, P. & Olsson, A. (2017). Endogenous opioids regulate social threat learning in humans. *Nat. Commun.*, *8*, 15495.

Hamm, A. O. & Vaitl, D. (1996). Affective learning: Awareness and aversion. *Psychophysiology*, *33*(6), 698–710.

Hamm, A. O., Vaitl, D. & Lang, P. J. (1989). Fear conditioning, meaning, and belongingness: A selective association analysis. *J. Abnorm. Psychol.*, *98*(4), 395–406.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362.

Hopwood, T. L. & Schutte, N. S. (2017). Psychological outcomes in reaction to media exposure to disasters and large-scale violence: A meta-analysis. *Psychol. Violence*, *7*(2), 316–327.

Jankowiak-Siuda, K., Duszyk, A., Dopierała, A., Bujwid, K., Rymarczyk, K. & Grabowska, A. (2019). Empathic responses for pain in facial muscles are modulated by actor's attractiveness and gender, and perspective taken by observer. *Front. Psychol.*, *10*, 624.

Jankowiak-Siuda, K., Rymarczyk, K., Żurawski, Ł., Jednoróg, K. & Marchewka, A. (2015). Physical attractiveness and sex as modulatory factors of empathic brain responses to pain. *Front. Behav. Neurosci.*, *9*, 236.

JASP Team. (2020). JASP.

Jenkinson, M., Bannister, P., Brady, M. & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825–841.

Jenkinson, M. & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.*, *5*(2), 143–156.

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S.-Y., Rabah, D., Kinet, J.-P. & Shin, H.-S. (2010). Observational fear learning involves affective pain system and cav1.2 ca2+ channels in ACC. *Nat. Neurosci.*, *13*(4), 482–488.

Jeon, D. & Shin, H.-S. (2011). A mouse model for observational fear learning and the empathetic response. *Curr. Protoc. Neurosci.*, *Chapter 8*, Unit 8.27.

Jolliffe, D. & Farrington, D. P. (2006). Development and validation of the basic empathy scale. *J. Adolesc.*, *29*(4), 589–611.

Kanwisher, N., McDermott, J. & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.*, *17*(11), 4302–4311.

Kaplan, D. & Rosenmann, A. (2014). Toward an empirical model of male homosocial relatedness: An investigation of friendship in uniform and beyond. *Psychol. Men Masc.*, *15*(1), 12–21.

Kassambara, A. (2020). *Ggpubr: 'ggplot2' based publication ready plots*.

Kassambara, A. (2021). *Rstatix: Pipe-Friendly framework for basic statistical tests*.

Kaźmierczak, M., Plopa, M. & Retowski, S. (2007). Skala wrażliwości empatycznej. *Przegląd Psychologiczny*, *50*(1), 9–24.

Kaźmierowska, A. M., Szczepanik, M., Wypych, M., Droździel, D., Marchewka, A., Michałowski, J. M., Olsson, A. & Knapska, E. (2021). Learning about threat from friends and strangers is equally effective: An fMRI study on observational fear conditioning. *bioRxiv*. https://doi.org/10.1101/2021.09.20.461036

Keysers, C. & Gazzola, V. (2009). Expanding the mirror: Vicarious activity for actions, emotions, and sensations. *Curr. Opin. Neurobiol.*, *19*(6), 666–671.

Keysers, C., Kaas, J. H. & Gazzola, V. (2010). Somatosensation in social perception. *Nat. Rev. Neurosci.*, *11*(6), 417–428.

Kikusui, T., Winslow, J. T. & Mori, Y. (2006). Social buffering: Relief from stress and anxiety. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *361*(1476), 2215–2228.

Knapska, E., Nikolaev, E., Boguszewski, P., Walasek, G., Blaszczyk, J., Kaczmarek, L. & Werka, T. (2006). Between-subject transfer of emotional information evokes specific pattern of amygdala activation. *Proc. Natl. Acad. Sci. U. S. A.*, *103*(10), 3858–3862.

Koch, M. (1999). The neurobiology of startle. *Prog. Neurobiol.*, *59*(2), 107–128.

Krabbe, S., Gründemann, J. & Lüthi, A. (2018). Amygdala inhibitory circuits regulate associative fear conditioning. *Biol. Psychiatry*, *83*(10), 800–809.

Kret, M. E. & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, *50*(7), 1211–1221.

Kuhn, M., Wendt, J., Sjouwerman, R., Büchel, C., Hamm, A. & Lonsdorf, T. B. (2020). The neurofunctional basis of affective startle modulation in humans: Evidence from combined facial electromyography and functional magnetic resonance imaging. *Biol. Psychiatry*, *87*(6), 548–558.

Laland, K. N. (2004). Social learning strategies. *Learn. Behav.*, *32*(1), 4–14.

Lamm, C., Decety, J. & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, *54*(3), 2492–2502.

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85.

LeDoux, J. E., Cicchetti, P., Xagoraris, A. & Romanski, L. M. (1990). The lateral amygdaloid nucleus: Sensory interface of the amygdala in fear conditioning. *J. Neurosci.*, *10*(4), 1062–1069.

LeDoux, J. E. (2003). The emotional brain, fear, and the amygdala. *Cell. Mol. Neurobiol.*, *23*(4-5), 727–738.

LeDoux, J. E. & Pine, D. S. (2016). Using neuroscience to help understand fear and anxiety: A Two-System framework. *AJP*, *173*(11), 1083–1093.

Lenth, R. (2021). *Emmeans: Estimated marginal means, aka Least-Squares means*.

Levy, D. P. (2005). Hegemonic complicity, friendship, and comradeship: Validation and causal processes among white, Middle-Class, Middle-Aged men. *The Journal of Men's Studies*, *13*(2), 199–224.

Lindström, B., Haaker, J. & Olsson, A. (2018). A common neural network differentially mediates direct and social fear learning. *Neuroimage*, *167*, 121–129.

Lissek, S., Pine, D. S. & Grillon, C. (2006). The strong situation: A potential impediment to studying the psychobiology and pharmacology of anxiety disorders. *Biol. Psychol.*, *72*(3), 265–270.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157.

Logothetis, N. K. & Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magn. Reson. Imaging*, *22*(10), 1517–1531.

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., … Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neurosci. Biobehav. Rev.*, *77*, 247–285.

Lovibond, P. F. & Shanks, D. R. (2002). The role of awareness in pavlovian conditioning: Empirical evidence and theoretical implications. *J. Exp. Psychol. Anim. Behav. Process.*, *28*(1), 3–26.

Maren, S. & Fanselow, M. S. (1996). The amygdala and fear conditioning: Has the nut been cracked? *Neuron*, *16*(2), 237–240.

Matusz, P. J., Dikker, S., Huth, A. G. & Perrodin, C. (2019). Are we ready for real-world neuroscience? *J. Cogn. Neurosci.*, *31*(3), 327–338.

McDiarmid, E., Gill, P. R., McLachlan, A. & Ali, L. (2017). "that whole macho male persona thing": The role of insults in young australian male friendships. *Psychol. Men Masc.*, *18*(4), 352–360.

Medford, N. & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: Awareness and response. *Brain Struct. Funct.*, *214*(5-6), 535–549.

Mendelson, M. J. & Aboud, F. E. (1999). Measuring friendship quality in late adolescents and young adults: McGill friendship questionnaires. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *31*(2), 130.

Mertens, G., Boddez, Y., Krypotos, A.-M. & Engelhard, I. M. (2021). Human fear conditioning is moderated by stimulus contingency instructions. *Biol. Psychol.*, *158*, 107994.

Mertens, G. & Engelhard, I. M. (2020). A systematic review and meta-analysis of the evidence for unaware fear conditioning. *Neurosci. Biobehav. Rev.*, *108*, 254–268.

Mineka, S., Davidson, M., Cook, M. & Keir, R. (1984). Observational conditioning of snake fear in rhesus monkeys. *J. Abnorm. Psychol.*, *93*(4), 355–372.

Morey, R. D. & Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs*.

Neisser, U. (1980). On "social knowing". *Pers. Soc. Psychol. Bull.*, *6*(4), 601–605.

Nichols, T. & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Stat. Methods Med. Res.*, *12*(5), 419–446.

Ojala, K. E. & Bach, D. R. (2020). Measuring learning in human classical threat conditioning: Translational, cognitive and methodological considerations. *Neurosci. Biobehav. Rev.*, *114*, 96–112.

Oliveira, R. F. & Faustino, A. I. (2017). Social information use in threat perception: Social buffering, contagion and facilitation of alarm responses. *Commun. Integr. Biol.*, *10*(3), e1325049.

Olsson, A., Knapska, E. & Lindström, B. (2020). The neural and computational systems of social learning. *Nat. Rev. Neurosci.*, *21*(4), 197–212.

Olsson, A., McMahon, K., Papenberg, G., Zaki, J., Bolger, N. & Ochsner, K. N. (2016). Vicarious fear learning depends on empathic appraisals and trait empathy. *Psychol. Sci.*, *27*(1), 25–33.

Olsson, A., Nearing, K. I. & Phelps, E. A. (2007). Learning fears by observing others: The neural systems of social fear transmission. *Soc. Cogn. Affect. Neurosci.*, *2*(1), 3–11.

Olsson, A. & Spring, V. (2018). Chapter 2 - the vicarious brain: Integrating empathy and emotional learning. In K. Z. Meyza & E. Knapska (Eds.), *Neuronal correlates of empathy* (pp. 7–23). Academic Press.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M. & Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. *Soc. Cogn. Affect. Neurosci.*, *7*(5), 604–609.

Pan, Y., Olsson, A. & Golkar, A. (2020). Social safety learning: Shared safety abolishes the recovery of learned threat. *Behav. Res. Ther.*, *135*, 103733.

Pärnamets, P., Espinosa, L. & Olsson, A. (2020). Physiological synchrony predicts observational threat learning in humans. *Proc. Biol. Sci.*, *287*(1927), 20192779.

Pavlov, I. P. (1928). *Lectures on conditioned reflexes: Twenty-five years of objective study of the higher nervous activity (behaviour) of animals*. Liverwright Publishing Corporation.

Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C. & Davis, M. (2001). Activation of the left amygdala to a cognitive representation of fear. *Nat. Neurosci.*, *4*(4), 437–441.

Phelps, E. A. & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*(2), 175–187.

Pineles, S. L., Orr, M. R. & Orr, S. P. (2009). An alternative scoring method for skin conductance responding in a differential fear conditioning paradigm with a long-duration conditioned stimulus. *Psychophysiology*, *46*(5), 984–995.

Poldrack, R. A., Mumford, J. A. & Nichols, T. E. (2011). *Handbook of functional MRI data analysis*. Cambridge University Press.

Poline, J.-B. & Brett, M. (2012). The general linear model and fMRI: Does love last forever? *Neuroimage*, *62*(2), 871–880.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L. & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665–678.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, *84*(Supplement C), 320–341.

Preston, S. D. & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behav. Brain Sci.*, *25*(1), 1–20.

Proverbio, A. M. (2021). Sex differences in the social brain and in social cognition. *J. Neurosci. Res.*

Purkis, H. M. & Lipp, O. V. (2001). Does affective learning exist in the absence of contingency awareness? *Learn. Motiv.*, *32*(1), 84–99.

Qi, Y., Herrmann, M. J., Bell, L., Fackler, A., Han, S., Deckert, J. & Hein, G. (2020). The mere physical presence of another person reduces human autonomic responses to aversive sounds. *Proc. Biol. Sci.*, *287*(1919), 20192241.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Reijmers, L. G., Perkins, B. L., Matsuo, N. & Mayford, M. (2007). Localization of a stable neural correlate of associative memory. *Science*, *317*(5842), 1230–1233.

Riečanský, I. & Lamm, C. (2019). The role of sensorimotor processes in pain empathy. *Brain Topogr.*, *32*(6), 965–976.

Robinson, S., White, A. & Anderson, E. (2019). Privileging the bromance: A critical appraisal of romantic and bromantic relationships. *Men Masc.*, *22*(5), 850–871.

Rütgen, M., Seidel, E.-M., Silani, G., Riečanský, I., Hummer, A., Windischberger, C., Petrovic, P. & Lamm, C. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proc. Natl. Acad. Sci. U. S. A.*, *112*(41), E5638–46.

Selbing, I. & Olsson, A. (2019). Anxious behaviour in a demonstrator affects observational learning. *Sci. Rep.*, *9*(1), 9181.

Sevenster, D., Beckers, T. & Kindt, M. (2014). Fear conditioning of SCR but not the startle reflex requires conscious discrimination of threat and safety. *Front. Behav. Neurosci.*, *8*, 32.

Silvers, J. A., Callaghan, B. L., VanTieghem, M., Choy, T., O'Sullivan, K. & Tottenham, N. (2021). An exploration of amygdala-prefrontal mechanisms in the intergenerational transmission of learned fear. *Dev. Sci.*, *24*(3), e13056.

Singmann, H., Bolker, B., Westfall, J., Aust, F. & Ben-Shachar, M. S. (2021). *Afex: Analysis of factorial experiments*.

Sjouwerman, R., Niehaus, J., Kuhn, M. & Lonsdorf, T. B. (2016). Don't startle me-interference of startle probe presentations and intermittent ratings with fear acquisition. *Psychophysiology*, *53*(12), 1889–1899.

Sliwinska, M. W. & Pitcher, D. (2018). TMS demonstrates that both right and left superior temporal sulci are important for facial expression recognition. *Neuroimage*, *183*, 394–400.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R. & Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U. S. A.*, *106*(31), 13040–13045.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R. & Jacobs, G. A. (1983). *Manual for the State-Trait anxiety inventory*. Consulting Psychologists Press.

Spielberger, C. D., Strelau, J., Tysarczyk, M. & Wrześniewski, K. (2012). *STAI - inwentarz stanu i cechy lęku*. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.

Staib, M., Castegnetti, G. & Bach, D. R. (2015). Optimising a model-based approach to inferring fear learning from skin conductance responses.

Szczepanik, M., Kaźmierowska, A. M., Michałowski, J. M., Wypych, M., Olsson, A. & Knapska, E. (2020). Observational learning of fear in real time procedure. *Sci. Rep.*, *10*(1), 16960.

Tange, O. (2011). GNU parallel - the Command-Line power tool. *;login: The USENIX Magazine*, *36*(1), 42–47.

Titchener, E. B. (1909). *Lectures on the experimental psychology of the thought-processes*. The Macmillan Company.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging*, *29*(6), 1310–1320.

Vack, N., Ollinger, J., Fitch, D. & Davison, M. (2019). Uwmadison-chm/bioread: 1.0.4.

van Heukelum, S., Mars, R. B., Guthrie, M., Buitelaar, J. K., Beckmann, C. F., Tiesinga, P. H. E., Vogt, B. A., Glennon, J. C. & Havenith, M. N. (2020). Where is cingulate cortex? a Cross-Species view. *Trends Neurosci.*, *43*(5), 285–299.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nat. Methods*, *17*(3), 261–272.

Wang, Y., Song, J., Guo, F., Zhang, Z., Yuan, S. & Cacioppo, S. (2016). Spatiotemporal brain dynamics of empathy for pain and happiness in friendship. *Front. Behav. Neurosci.*, *10*, 45.

Watson, J. B. & Rayner, R. (1920). Conditioned emotional reactions. *J. Exp. Psychol.*, *3*(1), 1–14.

Weidemann, G., Satkunarajah, M. & Lovibond, P. F. (2016). I think, therefore eyeblink: The importance of contingency awareness in conditioning. *Psychol. Sci.*, *27*(4), 467–475.

Weike, A. I., Schupp, H. T. & Hamm, A. O. (2008). In dubio pro defensio: Initial activation of conditioned fear is not cue specific. *Behav. Neurosci.*, *122*(3), 685–696.

Williams, A. L. & Conway, C. C. (2020). Empathy does not amplify vicarious threat learning. *Behav. Res. Ther.*, *127*, 103577.

Wittig, R. M., Crockford, C., Weltring, A., Langergraber, K. E., Deschner, T. & Zuberbühler, K. (2016). Social support reduces stress hormone levels in wild chimpanzees across stressful events and everyday affiliations. *Nat. Commun.*, *7*, 13361.

Yamamoto, S. (2017). Primate empathy: Three factors and their combinations for empathy-related phenomena. *Wiley Interdiscip. Rev. Cogn. Sci.*, *8*(3).

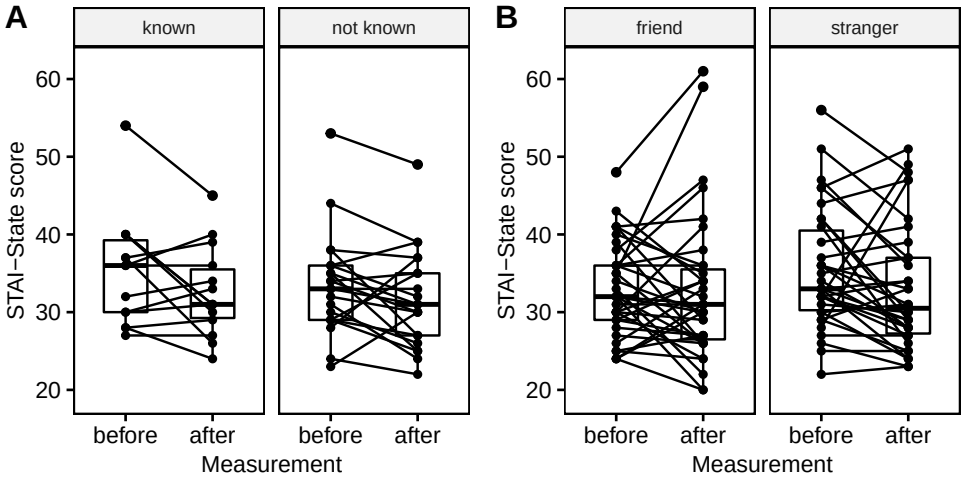
Yamamoto, S., Humle, T. & Tanaka, M. (2012). Chimpanzees' flexible targeted helping based on an understanding of conspecifics' goals. *Proc. Natl. Acad. Sci. U. S. A.*, *109*(9), 3588–3592.

Yang, D. Y.-J., Rosenblau, G., Keifer, C. & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neurosci. Biobehav. Rev.*, *51*, 263–275.

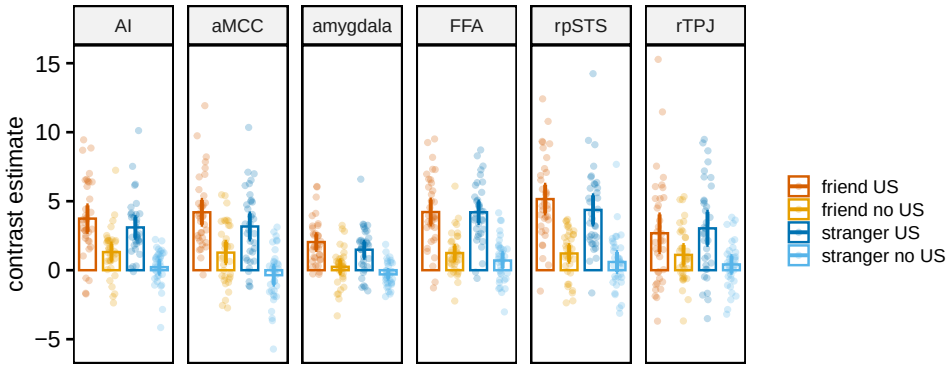
Zaki, J., Wager, T. D., Singer, T., Keysers, C. & Gazzola, V. (2016). The anatomy of suffering: Understanding the relationship between nociceptive and empathic pain. *Trends Cogn. Sci.*, *20*(4), 249–259.

Zhang, Y., Brady, M. & Smith, S. (2001). Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging*, *20*(1), 45–57.

Supplementary Figures



Supplementary Figure 1: Changes in STAI-state during the experiment. Observers completed the questionnaire before and after the experiment. (A) Results in the psychophysiological experiment. In ANOVA, there was a significant main effect of the measurement, and scores were higher before the experiment. (B) Neuroimaging experiment. In ANOVA, there were no significant effects. Joined dots represent individual subjects. The lower and upper hinges correspond to the first and third quartiles, and whiskers extend to the smallest (lower) or largest (upper) value no further than 1.5 × IQR from the hinge.



Supplementary Figure 2: Parameter estimates and 95% confidence intervals of US and no US responses. Parameter estimates were averaged for selected regions of interest. Dots represent individual subjects. AI – anterior insula, aMCC – anterior mid-cingulate cortex, Amy – amygdala, FFA – fusiform face area, rpSTS – right posterior superior temporal sulcus, rTPJ – right temporoparietal junction.

Supplementary Tables

Supplementary Table 1: Region of interest analysis for the observational learning stage, comparing US parameter estimates between friend and stranger groups. AI – anterior insula, aMCC – anterior mid-cingulate cortex, Amy – amygdala, FFA – fusiform face area, rpSTS – right posterior superior temporal sulcus, rTPJ – right temporoparietal junction.

ROI	t	p	BF_{01}
AI	1.06	.29	2.50
aMCC	1.60	.11	1.36
Amy	1.38	.17	1.79
FFA	0.03	.98	4.04
rpSTS	1.08	.29	2.47
rTPJ	−0.49	.69	3.78

Supplementary Table 2: Activation peaks in the observational learning stage, US > noUS contrast. Table shows local maxima more than 16 mm apart. For brevity, clusters greater than 30 voxels were included. Last column lists the most probable label from the Harvard - Oxford atlas. x, y, z - MNI space peak coordinates in MNI space, vox - number of voxels in the cluster

cluster	x	y	z	t	vox	label
1	50	−66	6	14.98	11239	Right Lateral Occipital Cortex inferior division
	44	−76	−8	12.92		Right Lateral Occipital Cortex inferior division
	42	−44	−16	12.21		Right Temporal Occipital Fusiform Cortex
	64	−40	26	11.94		Right Supramarginal Gyrus posterior division
	52	−28	−6	11.74		Right Middle Temporal Gyrus posterior division
	50	−40	10	11.51		Right Supramarginal Gyrus posterior division
	34	22	−18	10.71		Right Frontal Orbital Cortex
	52	8	−22	10.34		Right Temporal Pole
	22	−88	−8	10.07		Right Occipital Fusiform Gyrus
	46	26	−4	9.82		Right Frontal Orbital Cortex
	42	−2	−14	9.71		Right Planum Polare
	52	24	22	9.14		Right Inferior Frontal Gyrus pars opercularis
	46	2	44	8.88		Right Precentral Gyrus
	12	−68	38	8.81		Right Precuneous Cortex
	36	−62	−20	8.55		Right Occipital Fusiform Gyrus
	22	−2	−18	8.42		Right Amygdala
	14	−98	6	8.17		Right Occipital Pole
	28	−92	12	8.16		Right Occipital Pole
	38	6	4	7.50		Right Insular Cortex
	54	40	−2	7.40		Right Frontal Pole
	64	−32	40	7.35		Right Supramarginal Gyrus anterior division
	26	−80	34	6.78		Right Lateral Occipital Cortex superior division
2	−48	−78	6	13.87	4071	Left Lateral Occipital Cortex inferior division
	−52	−58	10	11.73		Left Middle Temporal Gyrus temporooccipital part
	−62	−46	22	11.32		Left Supramarginal Gyrus posterior division
	−44	−50	−18	10.52		Left Inferior Temporal Gyrus temporooccipital part
	−30	−92	−10	9.49		Left Occipital Pole
	−24	−66	−28	6.64		no label
	−28	−94	14	6.22		Left Occipital Pole
3	−32	20	−14	10.65	1940	Left Frontal Orbital Cortex
	−30	28	2	8.87		Left Insular Cortex
	−42	0	−16	8.24		Left Planum Polare
	−38	−12	−6	8.12		Left Insular Cortex
	−20	−4	−14	7.58		Left Amygdala
	−54	10	8	6.42		Left Inferior Frontal Gyrus pars opercularis
4	6	20	38	9.50	1296	Right Paracingulate Gyrus
	4	20	56	9.07		Right Superior Frontal Gyrus
	6	40	10	8.55		Right Cingulate Gyrus anterior division
	10	4	72	6.61		Right Superior Frontal Gyrus

Continued on next page

Supplementary Table 2: Activation peaks in the observational learning stage, US > noUS contrast.

cluster	x	y	z	t	vox	label
5	−8	−24	−12	9.77	852	Brain-Stem
	12	−22	−10	8.72		no label
	10	−8	6	8.38		Right Thalamus
	10	8	4	7.85		Right Caudate
	6	−4	−10	7.13		no label
6	−10	−66	36	8.61	292	Left Precuneous Cortex
7	−2	−24	28	8.69	262	Left Cingulate Gyrus posterior division
	0	−8	32	6.43		Left Cingulate Gyrus anterior division
8	−16	−94	26	7.12	237	Left Occipital Pole
	−26	−78	24	5.95		Left Lateral Occipital Cortex superior division
9	34	−52	50	6.78	134	Right Superior Parietal Lobule
10	−30	−52	52	6.86	85	Left Superior Parietal Lobule
11	4	52	34	6.47	79	Right Superior Frontal Gyrus
12	−10	−76	−44	6.58	55	no label
13	−10	−84	8	6.03	49	Left Intracalcarine Cortex
14	−40	−4	48	6.68	47	Left Precentral Gyrus
15	−50	−26	−6	7.12	46	Left Middle Temporal Gyrus posterior division

Supplementary Table 3: Activation peaks in the direct-expression stage, CS+ > CS– contrast. Table shows local maxima more than 16 mm apart. Last column lists the most probable label from the Harvard - Oxford atlas. x, y, z - MNI space peak coordinates in MNI space, vox - number of voxels in the cluster

cluster	x	y	z	t	vox	label
1	30	30	0	8.30	628	Right Frontal Orbital Cortex
	50	22	4	6.70		Right Inferior Frontal Gyrus pars triangularis
2	−36	24	−4	7.18	175	Left Frontal Orbital Cortex
3	−14	−76	−30	6.37	14	no label
4	10	−12	−12	6.25	8	no label
5	6	24	36	5.77	5	Right Paracingulate Gyrus
6	34	46	20	5.63	1	Right Frontal Pole
7	8	4	2	5.45	1	Right Caudate

Supplementary Table 4: Activation peaks for the temporal modulation of the CS+ response in the direct-expression stage. Table shows local maxima more than 16 mm apart. Last column lists the most probable label from the Harvard - Oxford atlas. x, y, z - MNI space peak coordinates in MNI space, vox - number of voxels in the cluster

cluster	x	y	z	t	vox	label
				Both groups		
1	−18	−88	−6	6.38	1310	Left Occipital Fusiform Gyrus
	−8	−100	0	5.43		Left Occipital Pole
	−12	−80	−32	4.49		no label
	10	−86	−10	4.47		Right Lingual Gyrus
	0	−86	4	4.01		Left Intracalcarine Cortex
	24	−76	−14	3.65		Right Occipital Fusiform Gyrus
	−26	−98	20	3.58		Left Occipital Pole
2	−4	−54	62	5.42	668	Left Precuneous Cortex
	−12	−76	44	4.84		Left Lateral Occipital Cortex superior division
	4	−36	46	3.75		Right Cingulate Gyrus posterior division
	−8	−68	28	3.71		Left Precuneous Cortex
3	2	8	34	4.58	213	Right Cingulate Gyrus anterior division
	−2	26	24	3.66		Left Cingulate Gyrus anterior division
4	44	28	8	4.71	148	Right Inferior Frontal Gyrus pars triangularis
5	6	46	32	5.67	106	Right Paracingulate Gyrus
				Friend > Stranger		
1	−58	14	10	5.83	285	Left Inferior Frontal Gyrus pars opercularis
2	−62	−44	8	4.65	166	Left Supramarginal Gyrus posterior division
3	16	−58	−2	4.41	137	Right Lingual Gyrus
4	−52	36	8	4.83	134	Left Inferior Frontal Gyrus pars triangularis
5	−24	−6	12	4.55	96	Left Putamen

Supplementary Table 5: Activation peaks for the psychophysiological interaction analysis based on the US > no US contrast. Table shows local maxima more than 16 mm apart. Last column lists the most probable label from the Harvard - Oxford atlas. x, y, z - MNI space peak coordinates in MNI space, vox - number of voxels in the cluster

cluster	x	y	z	t	vox	label
				Anterior insula		
1	64	−38	20	6.16	1065	Right Supramarginal Gyrus posterior division
	54	−48	4	5.00		Right Middle Temporal Gyrus temporooccipital part
	58	−66	6	4.63		Right Lateral Occipital Cortex inferior division
	48	−28	2	4.62		Right Superior Temporal Gyrus posterior division
	48	−80	0	3.60		Right Lateral Occipital Cortex inferior division
2	−52	−56	12	5.26	849	Left Middle Temporal Gyrus temporooccipital part
	−42	−70	12	4.80		Left Lateral Occipital Cortex inferior division
	−68	−46	8	4.01		Left Middle Temporal Gyrus temporooccipital part
	−50	−80	−4	3.76		Left Lateral Occipital Cortex inferior division
3	−12	−74	−14	4.88	442	Left Lingual Gyrus
	−30	−74	−22	4.64		Left Occipital Fusiform Gyrus
	−40	−64	−14	3.66		Left Occipital Fusiform Gyrus
4	52	14	26	4.46	101	Right Inferior Frontal Gyrus pars opercularis
5	32	−68	−28	4.16	93	no label
				Posterior STS		
1	46	−64	2	9.29	2582	Right Lateral Occipital Cortex inferior division
	66	−40	20	6.51		Right Supramarginal Gyrus posterior division
	38	−80	−14	5.01		Right Lateral Occipital Cortex inferior division
	46	−40	10	4.64		Right Supramarginal Gyrus posterior division
	62	−38	44	4.21		Right Supramarginal Gyrus posterior division
	32	−94	−4	3.51		Right Occipital Pole
2	−46	−70	8	8.07	1553	Left Lateral Occipital Cortex inferior division
	−58	−48	12	5.20		Left Supramarginal Gyrus posterior division
3	38	32	−2	5.68	605	Right Frontal Orbital Cortex
4	−58	−40	36	4.70	384	Left Supramarginal Gyrus posterior division
	−66	−32	24	4.63		Left Supramarginal Gyrus anterior division
5	−38	22	−8	5.01	326	Left Frontal Orbital Cortex
	−34	6	−6	3.57		Left Insular Cortex
6	48	14	30	4.91	280	Right Inferior Frontal Gyrus pars opercularis
	54	30	20	3.71		Right Inferior Frontal Gyrus pars triangularis
7	48	−42	−16	6.12	249	Right Inferior Temporal Gyrus temporooccipital ...
8	38	−12	−6	4.76	221	Right Insular Cortex
	50	−24	−4	4.49		Right Middle Temporal Gyrus posterior division
9	12	−2	12	4.93	208	Right Thalamus
	−2	−4	4	3.94		Left Thalamus
10	54	20	8	4.96	190	Right Inferior Frontal Gyrus pars opercularis
11	6	6	62	4.63	168	Right Juxtapositional Lobule Cortex (formerly S...
12	20	10	−6	4.96	140	Right Putamen

Continued on next page

Supplementary Table 5: Activation peaks for the psychophysiological interaction analysis based on the US > no US contrast.

cluster	x	y	z	t	vox	label
13	−4	−14	−10	4.75	113	no label
14	6	−24	0	4.49	102	Right Thalamus
15	−40	−62	−28	4.65	93	no label
Posterior STS (small volume corrected within the amygdala)						
1	24	−4	−16	3.82	13	Right Amygdala
2	−22	−4	−14	3.56	3	Left Amygdala
3	−20	−6	−18	3.42	1	Left Amygdala

Own publications

Authors who equally contributed to a publication are marked with a †.

Observational fear conditioning

1. Kaźmierowska, A. M.†, **Szczepanik, M.**†, Wypych, M., Droździel, D., Marchewka, A., Michałowski, J. M., Olsson, A., & Knapska, E. (2021). Learning about threat from friends and strangers is equally effective: an fMRI study on observational fear conditioning [preprint]. In *bioRxiv*. https://doi.org/10.1101/2021.09.20.461036

2. **Szczepanik, M.**†, Kaźmierowska, A. M.†, Michałowski, J. M., Wypych, M., Olsson, A., & Knapska, E. (2020). Observational learning of fear in real time procedure. *Scientific Reports, 10(1)*, 16960.

Other

3. Banaszkiewicz, A., Bola, Ł., Matuszewski, J., **Szczepanik, M.**, Kossowski, B., Mostowski, P., Rutkowski, P., Śliwińska, M., Jednoróg, K., Emmorey, K., & Marchewka, A. (2021). The role of the superior parietal lobule in lexical processing of sign language: Insights from fMRI and TMS. *Cortex, 135*, 250–254.

4. Kuper, C., Matuszewski, J., Banaszkiewicz, A., Bola Ł., **Szczepanik, M.**, Draps, M., Kordos, P., Szwed, M., Jednoróg, K., & Marchewka, A. (2021). Functional reorganization of the reading network in the course of foreign language acquisition. *Neuroimage, 226*, 117544.

5. Banaszkiewicz, A., Matuszewski, J., Bola, Ł., **Szczepanik, M.**, Kossowski, B., Rutkowski, P., Szwed, M., Emmorey, K., Jednoróg, K., & Marchewka, A. (2021). Multimodal imaging of brain reorganization in hearing late learners of sign language. *Human Brain Mapping, 42(2)*, 384-397.

6. Bola, Ł., Matuszewski, J., **Szczepanik, M.**, Droździel, D., Śliwinska, M. W., Paplińska, M., Jednoróg, K., Szwed, M., & Marchewka, A. (2019). Functional hierarchy for tactile processing in the visual cortex of sighted adults. *NeuroImage, 202*, 116084.

7. Wypych, M., Michałowski, J. M., Droździel, D., Borczykowska, M., **Szczepanik, M.**, & Marchewka, A. (2019). Attenuated brain activity during error processing and punishment anticipation in procrastination – a monetary Go/No-go fMRI study. *Scientific Reports, 9(1)*, 11492.