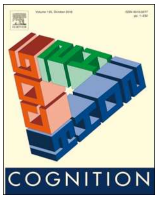
Cognition xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit



Individual differences in social and non-social cognitive control

Kohinoor M. Darda[a,b,c], Emily E. Butler[a], Richard Ramsey[d,*]

[a] *Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, Gwynedd, Wales LL57 2AS, United Kingdom*
[b] *Institute of Neuroscience & Psychology, University of Glasgow, UK*
[c] *Department of Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia*
[d] *Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia*

ARTICLE INFO

Keywords:
Cognitive control
Sex differences
Personality
Automatic imitation

ABSTRACT

Cognitive control refers to the ability of human beings to adapt flexibly and quickly to continuously changing environments. Several decades of research have identified a diverse range of mental processes that are associated with cognitive control but the extent to which shared systems underlie cognitive control in social and non-social contexts, as well as how these systems may vary across individuals, remains largely unexplored. By integrating methodological approaches from experimental and differential psychology, the current study is able to shine new light on the relationships between stable features of individuals, such as personality and sex, and the architecture of cognitive control systems using paradigms that index social (automatic imitation) and spatial processes. Across three large-sample experiments (> 600 participants in total), we demonstrate that cognitive control systems are largely invariant to stable aspects of personality, but exhibit a sex difference, such that females show greater task-interference than males. Moreover, we further qualified this sex difference in two ways. First, we showed that the sex difference was unrelated to the sex of the interaction partner and therefore did not reflect an in-group bias based on sex. Second, we showed that the sex difference was tied to a form of spatial interference control rather than social (imitative) control and therefore it does not reflect a specialised mechanism for guiding social interactions exclusively. Instead, our findings suggest that a robust sex difference exists in the system (or set of subsystems) that operate in resolving a form of spatial interference control, and that such systems are unaffected by social factors such as the sex of the interaction partner. The results highlight the value of integrating approaches from experimental and differential psychology by providing a deeper understanding of the structure of cognitive control systems, while also providing new dimensions to incorporate into theories and models of social and non-social control.

1. General introduction

A remarkable feature of the human cognitive system is its ability to quickly and flexibly adapt behaviour to guide interactions with people and objects in the environment. The mental processes behind such adaptability are collectively referred to as cognitive control and have been the focus of growing research in cognitive psychology and cognitive neuroscience (Banich, 2009; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Botvinick & Cohen, 2014; Inzlicht, Bartholow, & Hirsh, 2015; O'Reilly, Herd, & Pauli, 2010). To date, however, the extent to which shared systems underlie cognitive control in social and non-social contexts, as well as how these systems may vary across individuals, remains largely unexplored. As such, for a more complete understanding of the structure of social and non-social cognition, we need to investigate how general cognitive mechanisms operate across all individuals and contexts, as well as how different individuals vary from these general patterns (de Schotten & Shallice, 2017; Fischer-Baum, Kook, Lee, Ramos Nuñez, & Vannucci, 2018). Thus, in the current paper, across three experiments, we integrate methodological approaches from experimental and differential psychology to investigate the extent to which cognitive systems relating to social (imitative) and non-social control differ between individuals, and whether such individual differences rely on domain-general or specialised control mechanisms.

Cognitive control is multi-faceted, with a core function being the ability to inhibit unwanted but dominant responses, in order to prioritise alternative, context-appropriate responses (Chaiken & Trope, 1999; Miyake et al., 2000; Payne, 2005). For example, non-social cognitive control may involve inhibiting automatic reading responses in a Stroop task (MacLeod, 1991), whereas social cognitive control may

---

* Corresponding author.
E-mail addresses: kohinoor.darda@mq.edu.au (K.M. Darda), richard.ramsey@mq.edu.au (R. Ramsey).

https://doi.org/10.1016/j.cognition.2020.104317
Received 2 October 2019; Received in revised form 28 April 2020; Accepted 28 April 2020

Please cite this article as: Kohinoor M. Darda, Emily E. Butler and Richard Ramsey, Cognition, https://doi.org/10.1016/j.cognition.2020.104317

involve controlling automatic social biases based on race, sex or other social groupings (Amodio et al., 2004). The study of cognitive processes during social interactions has received much attention in the last couple of decades across different methodologies (Adolphs, 2009; Frith, 2008; Frith & Frith, 2012; Ochsner & Lieberman, 2001). For example, researchers investigating social cognition have used methodologies from cognitive psychology to provide experimental control over phenomena of interest to social psychologists (Lambert & Scherer, 2013). One such example is that of methodologies used to study automatic imitation.

Humans imitate a wide range of behaviours from their interaction partners, including speech patterns, body postures, gestures and facial expressions (Bernieri, 1988; Brass, Bekkering, Wohlschlager, & Prinz, 2000; Dimberg, 1982; Hansen, Alves, & Trope, 2016; Webb, 1972). This behaviour is usually not intended, often occurs without the conscious awareness of the imitator, and is termed automatic imitation (Heyes, 2009, 2011). Automatic imitation has been argued to function as a "social glue," powering cognitive and social development, enhancing emotional reciprocity, and increasing feelings of affiliation, positive rapport and pro-social behaviour (Cacioppo, Berntson, Sheridan, & McClintock, 2000; Chartrand & Bargh, 1999; Kavanagh & Winkielman, 2016; Lakin & Chartrand, 2003; Van Baaren, Holland, Steenaert, & van Knippenberg, 2003; van Baaren, Janssen, Chartrand, & Dijksterhuis, 2009). Given the substantial role automatic imitation plays in our social world, it has received attention from diverse research disciplines. For instance, in social psychology, studies of automatic imitation (termed motor mimicry) have typically involved measuring overt copying behaviours during live social interactions (Chartrand & Bargh, 1999; Ray & Heyes, 2011). In one such study, Chartrand and Bargh (1999) showed that even though it was unrelated to the experimental task, participants copied a confederate's behaviour such that they touched their face more than waggled their foot when the confederate also touched their face, and vice versa.

By contrast, in cognitive psychology, stimulus-response compatibility (SRC) paradigms have been used in order to measure automatic imitation (Brass et al., 2000; Kilner, Paulignan, & Blakemore, 2003; Stürmer, Aschersleben, & Prinz, 2000). In a typical SRC task measuring automatic imitation, participants are required to lift their index or middle finger in response to a number cue ('1' for index finger, '2' for middle finger). Simultaneously, they see either the same finger movement (compatible condition) or a different finger movement (incompatible condition). Participants respond slower in the incompatible condition as the observed movement interferes with their response. This difference in reaction time between compatible and incompatible conditions is referred to as the compatibility effect and is considered to be a measure of automatic imitation control. The logic of the task is based on the idea that the observed action generates a motor representation of the same action in the observer. Thus, it has been suggested that on incompatible trials cognitive resources are required to inhibit the automatic tendency to copy an observed (incorrect) action and instead prioritise the alternative (correct) action (Brass & Heyes, 2005).

After establishing the basic SRC paradigm in order to measure a form of social (imitative) control, subsequent research on automatic imitation in cognitive psychology has provided insight into the factors that influence automatic imitation (Cracco et al., 2018; Heyes, 2011). These studies have typically used an experimental method, which measures the average influence of a manipulation across a group of participants, rather than a differential approach that measures differences across individuals. For example, previous research has found that factors like eye gaze and facial expressions of the interacting partner modulate the tendency to automatically imitate (Wang, Ramsey, & Hamilton, 2011; Wang & Hamilton, 2014; Grecucci et al., 2011; Crescentini, Mengotti, Grecucci, & Rumiati, 2011; Rauchbauer, Majdandžić, Hummer, Windischberger, & Lamm, 2015, Butler, Ward, R., & Ramsey, 2016). These findings suggest that social and contextual factors serve as antecedents to automatic imitative behaviours.

Although a confluence of experimental and differential approaches has been suggested as a step toward unification and to aid progress of psychological science as a whole (Cronbach, 1975; Eysenck, 1997; Eysenck & Eysenck, 1985), these two streams of thought have remained largely autonomous (Cronbach, 1957; Cramer, Waldorp, Van Der Maas, & Borsboom, 2010). By focussing on the experimental method, the contribution of individual differences tends to be neglected (Eysenck, 1997). For example, in the context of social information processing, a recent study found that tasks measuring mental state reasoning may reflect socioeconomic characteristics of the sample as much as socio-cognitive processes (Dodell-Feder, Ressler, & Germine, 2019). Thus, it is essential to embrace both experimental and differential methods (including but not limited to sex, age, social class, culture, and personality traits) in order to fully understand the complex underpinnings of social interactions.

To aid cross-pollination between experimental and differential approaches, more recent imitation research has started to take an individual differences approach by investigating how characteristics of the imitator such as empathy, narcissism, alexithymia and interoceptive awareness, influence automatic imitation (Ainley, Brass, & Tsakiris, 2014; Chartrand & Bargh, 1999; Hogeveen & Obhi, 2013; Obhi, Hogeveen, Giacomin, & Jordan, 2013; Sowden, Brewer, Catmur, & Bird, 2016). Such claims, however, are limited due to the small number of studies reported to date, together with the use of relatively small sample sizes and a lack of powerful replications. Moreover, further studies, which used considerably larger sample sizes, have not been able to replicate the moderating influence of personality variables on automatic imitation (Butler, Ward, & Ramsey, 2015; Cracco et al., 2018). Interestingly, however, Butler et al. (2015) showed that the sex of the participant modulated the compatibility effect such that females showed a greater compatibility effect compared to males. Therefore, it is possible that biological sex is a factor to consider further when attempting to understand how cognitive mechanisms supporting imitation vary across individuals.

Sex is an important individual difference that influences a wide range of cognitive abilities and skills (Geary, 2010), as well as sensitivity to non-verbal social cues (Hall, 1978). However, few studies have investigated how socio-cognitive abilities vary as a function of biological sex and the ones that do have typically focused on mental reasoning or emotion perception (Campbell et al., 2002; Krach et al., 2009; Rahman, Wilson, & Abrahams, 2004; Russell, Tchanturia, Rahman, & Schmidt, 2007). Further, such prior studies have often produced mixed results based on relatively small sample sizes (Hyde, 2014; Miller & Halpern, 2014). Therefore, the potential influence of sex on complex cognitive mechanisms that control non-verbal interactions, remains largely unknown.

The extent to which sex differences operate in imitative behaviour has also received minimal attention to date. For example, no sex differences have been found on the automatic imitation of actions or gestures (Chartrand & Bargh, 1999; Larsen, Overbeek, Granic, & Engels, 2010). By contrast, studies on facial mimicry have shown that females automatically imitate facial expressions more than males (Dimberg, 1990; Sonnby-Borgström, Jönsson, & Svensson, 2008). Although there is no consistent empirical evidence to date, which speaks to sex differences in imitation, there is theoretical reason to think that sex differences may exist in imitative behaviour. Indeed, there is robust evidence for females to be more empathetic than males (Baron-Cohen & Wheelwright, 2004; Christov-Moore et al., 2014). Further, empathy has been associated with a variety of paradigms investigating imitation (Chartrand & Bargh, 1999; Müller, Leeuwen, Baaren, Bekkering, & Dijksterhuis, 2013; Sonnby-Borgström, 2002). Therefore, given that females are likely to be more empathetic than males in general, this may lead to more imitative tendencies in particular social contexts.

A core question pertains to whether the sex difference seen in the SRC task is a genuine difference between males and females or reflects an in-group/own-sex bias. The stimuli used in the task by Butler et al.

(2015) were of a female hand. It may be that automatic imitation increases when the sex of the participant and the interacting partner are matched. Group biases (typically in-group favouritism and out-group dislike) are prevalent in day-to-day social interactions (Allport, 1954; Cameron, Alvarez, Ruble, & Fuligni, 2001), relating to race, ethnicity (Ito & Bartholow, 2009; Kubota, Banaji, & Phelps, 2012; Malpass & Kravitz, 1969; van Bavel & Cunningham, 2009; Milner, 1983; Aboud, 1988), sex (Brown, 1995; Fishbein, 1996; Powlishta, 1995; Rudman & Goodwin, 2004; Yee & Brown, 1994), and arbitrary groups (Bernstein, Young, & Hugenberg, 2007; Tajfel, Billig, Bundy, & Flament, 1971). Therefore, ingroup biases seem like a powerful mechanism, which may guide imitative behaviour based on the sex of the interaction partner. In imitation research specifically, children have been known to imitate same-sex models more than others (Shutts, Banaji, & Spelke, 2010). Facial imitation and SRC measures of automatic imitation have both been found to increase when the interacting partner is an in-group member compared to an out-group member based on race, ethnicity and arbitrary group assignment (Gleibs, Wilson, Reddy, & Catmur, 2016; Mondillon, Niedenthal, Gil, & Droit-Volet, 2007; Rauchbauer et al., 2015). Moreover, recent work provides suggestive evidence for a sex difference and/or in-group bias in the automatic imitation task (Cracco et al., 2018; Genschow et al., 2017). For example, a meta-analysis found a higher reaction time compatibility effect when the sex of the stimuli matched the sex of the majority of participants in the sample (Cracco et al., 2018).[1] However, the extent to which this sex difference reflects an in-group bias remains unclear because no existing study has manipulated the sex of the stimuli across male and female participants.

Two other possible explanations exist. First, the sex difference on the automatic imitation SRC task could reflect that females tend to automatically imitate more than males, and therefore require more cognitive resources to inhibit the tendency to automatically imitate, leading to a greater compatibility effect. If so, the sex difference would be tied to a process related to imitation specifically. Second, the sex difference may be more domain-general in nature i.e. it may reflect a basic difference in the cognitive systems that underlie performance on SRC tasks more generally. Consistent with a domain-general explanation, sex differences have been found on many non-social inhibitory control tasks which, like the imitation task, require the inhibition of task-irrelevant automatic response tendencies in order to enforce a task-relevant response (e.g. flanker, oddball, gaze- and arrow-cueing, and Simon tasks; Stoet, 2010; Judge & Taylor, 2012; Clayson, Clawson, & Larson, 2011; Rubia, Hyde, Halari, Giampietro, & Smith, 2010; Bayliss, di Pellegrino, & Tipper, 2005; Merritt et al., 2007; Alwall, Johansson, & Hansen, 2010). As such, sex differences in SRC tasks may reflect differences in cognitive systems that operate across these tasks such as selective attention (Clayson et al., 2011) and/or spatial processing (Stoet, 2017).

These findings suggest that it is as yet unclear whether the sex difference on SRC measures of automatic imitation reflect more domain-general processes or processes solely tied to imitative control (Butler et al., 2015; Cracco et al., 2018; Darda, Butler, & Ramsey, 2018; Darda & Ramsey, 2019; Genschow et al., 2017). The SRC task used by researchers to demonstrate the existence of a sex difference was a composite of both spatial and imitative components (Butler et al., 2015; Genschow et al., 2017). A sex difference solely tied to imitative control might suggest that a distinct mechanism, or a partially distinct set of mechanisms, may underpin performance on the automatic imitation task compared to other inhibitory control tasks. Thus, in order to understand the cognitive architecture of social interactions, it is critical to unpack the relative contributions of both general and specific components in socio-cognitive processes (Michael & D'Ausilio, 2015; Spunt & Adolphs, 2017; Binney & Ramsey, 2019; Ramsey, 2018; Ramsey & Ward, 2020). Therefore, in the current study, we investigate sex differences on the automatic imitation task as well as a non-social control task in order to investigate whether the sex difference relies on domain-general and/or domain-specific mechanisms.

In the current paper, across three large-sample experiments, we integrate approaches from experimental and differential psychology approaches to investigate critical questions pertaining to individual differences in a form of social (imitative) and non-social cognitive control. First, consistent with recent suggestions to make replication a common and foundational practice in psychology (Zwaan, Etz, Lucas, & Donnellan, 2018), we aim to confirm the sex difference found previously (on both social and non-social cognitive control tasks) and provide a more precise estimate of the effect size. Further, we aimed to replicate the lack of variation in automatic imitation as a function of personality traits that has been reported previously in large sample research designs (Butler et al., 2015; Cracco et al., 2018). Second, we aim to investigate whether the sex difference on the imitation task reflects an actual difference between males and females, or an in-group or own-sex bias. Third, we aim to uncover whether mechanisms underlying the sex difference are domain-general or domain-specific (or a combination of both).

2. Experiment 1

2.1. Introduction

In the first experiment, we aim to replicate the sex difference on the general compatibility effect found previously (Butler et al., 2015; Genschow et al., 2017). We extend this research by investigating whether performance on a non-social inhibitory control task (the flanker task) also varies between the sexes. A similar sex difference on both tasks would indicate that the sex difference is supported by differences in a basic domain-general control system that underpin performance across social and non-social tasks. Alternatively, a sex difference on only one task would indicate at least partially distinct mechanisms as a function of sex.

Further, we also investigate the extent to which stable dimensions of personality influence the control of automatic imitation as measured on the SRC task. Prior work has provided mixed evidence regarding this question. Some studies have found a link between automatic imitation and empathy and narcissism - automatic imitation was higher for individuals who scored high on the empathy scale, and lower for those who scored higher on the narcissism scale (Chartrand & Bargh, 1999; Hogeveen & Obhi, 2013; Obhi et al., 2013). There are theoretical grounds to also posit a link between automatic imitation and two of the Big Five personality factors. Agreeableness and extraversion have been previously linked to empathy, altruism, and sociability (Ashton, Paunonen, Helmes, & Jackson, 1998; Barrio, Aluja, & García, 2004; McCrae & Costa Jr, 1999), and are thus considered as contributors to prosocial behaviour (Graziano & Eisenberg, 1997). Thus, individuals who are more agreeable and more extraverted may be more prosocial and could thus imitate their interacting partners more than others. In addition, although debated (Hamilton, 2013; Southgate & Hamilton, 2008), imitation abilities have been argued to vary in atypical populations including autism spectrum disorders and schizophrenia (Oberman & Ramachandran, 2007; Thakkar, Peterman, & Park, 2014; Williams, Whiten, Suddendorf, & Perrett, 2001), indicating that a relationship may exist between autistic-like and schizotypal traits and automatic imitation.

The largest datasets to date, however, show that performance on the SRC task is invariant to stable personality variables (Butler et al., 2015; Cracco et al., 2018). One concern with such null effects of personality is that they may reflect the impoverished social context of the SRC task.

---

[1] However, the authors of the meta-analysis categorised a sample as "female" if more than half the population was female. Thus, even samples with 51% females would be classified as a female sample, biasing the interpretation of the consequent analysis and making clear conclusions difficult to reach.

That is, effects of interest may only operate in more socially meaningful contexts. Therefore, in Experiment 1, we make the social context more meaningful by including emotional facial expressions within our design and investigate the extent to which automatic imitation continues to remain invariant as a function of personality. We included five face images depicting five emotional expressions (fearful, angry, happy, sad, neutral), and personality variables included extraversion, agreeableness, autistic-like and schizotypal traits, narcissism (grandiose and vulnerability narcissism), empathy (empathic concern and perspective taking), and alexithymia (for detailed information about measures used, see Supplementary material). We included only the perspective taking and empathic concern subscales of the empathy questionnaire as we had directional predictions about these subscales – prior evidence suggests a positive link between perspective taking and imitation, as well as empathic concern and imitation (Chartrand & Bargh, 1999; Santiesteban et al., 2012). Following Butler et al. (2015), although there is reason to expect pro-social dimensions of personality to be related to increased imitation, based on prior empirical evidence using the SRC task, we would expect imitation to be invariant to stable dimensions of personality.

2.2. *Method*

Across all experiments, we report how the sample size was determined, all data exclusions, and all measures in the study (Simmons, Nelson, & Simonsohn, 2011; 2012). Following open science initiatives, all raw data are available online for other researchers to pursue alternative questions of interest. For all three experiments, data pre-processing, statistical analyses, and data visualisations were performed using R (R Core Team, 2018), unless otherwise specified. All raw data and code used for analyses are available online (https://osf.io/fsh9b/). For all following experiments, we report our primary and secondary questions of interest in Table 1.

We determined the sample size for our experiments as follows. For experiment 1, we aimed to collect as many participants as possible over a two-day data collecting session. Therefore, the stopping rule was to terminate data collection after day 2 of data collection. For Experiments 2 and 3, in order to focus our design on the primary research question, which concerned sex differences, we set a minimum sample size of 100 male and 100 female participants. Sensitivity analyses revealed that given a sample size of two hundred participants (100 per sex), we would have 80% power to detect an effect size of Cohen's $d > 0.35$ for the mean difference between the two sexes on a one-tailed *t*-test, and an effect size of $\eta_p^2 > 0.04$ for a 2 × 2 mixed ANOVA. Such a design, therefore, provides reasonable confidence (80%) to detect effect sizes of interest that are conventionally considered small-to-medium.

2.2.1. *Participants*

Two hundred and three participants took part in this experiment for monetary compensation (£6) or course credit. All participants provided informed consent, had normal or corrected-to-normal vision, and were right-handed. Approval was obtained from the Research Ethics and Governance Committee of the School of Psychology at Bangor University. Participants were excluded if performance was three standard deviations away from the group mean average performance per condition in terms of accuracy or reaction time ($N = 14$ for the imitation task, $N = 7$ for the flanker task). A further 14 participants were excluded as demographic information (sex of the participant, or both age and sex of the participant) was not recorded. For the imitation task, the final sample included 175 participants (59 males, $\text{Mean}_{\text{age}} = 20.9$, $\text{SD}_{\text{age}} = 4.23$). For the flanker task, the final sample included 182 participants (59 males, $\text{Mean}_{\text{age}} = 20.9$, $\text{SD}_{\text{age}} = 3.73$).

2.2.2. *Stimuli, tasks, and procedure*

2.2.2.1. Automatic imitation task. The automatic imitation task was based on the stimulus response compatibility (SRC) paradigm developed by Brass et al. (2000), which consisted of the observation and execution of finger lifting movements (Fig. 1). In order to explore whether facial cues signalling emotional states influenced automatic imitation, five face images depicting five different emotional states were also presented along with the hand stimuli of the imitation task. The face stimuli were images of 5 female individuals from the NimStim data set with five different expressions (neutral, sad, happy, fearful, and angry) (Tottenham et al., 2009). The hand stimuli comprised five images of a female hand positioned in the centre of the screen and viewed from a third person perspective such that the fingers extended toward the participants. The first image was of the hand in a neutral position, while the remaining four images showed either an index or middle finger lift with a number ‘1’ or ‘2’ presented between the index and middle finger.

Participants were asked to hold down the “m” and “n” keys on the keyboard with their index and middle fingers of the right hand, respectively. They were instructed to lift their index finger when they saw a number “1” and their middle finger when they saw the number “2”. Thus, there were four possible trial types, two of which were compatible, and two of which were incompatible. In the compatible condition, participants were cued to perform the same finger-lifting movement that they observed (i.e. an index finger movement with a ‘1’ or a middle finger movement with a ‘2’). In the incompatible condition, the executed and observed movements were different (i.e. an index finger movement with a ‘2’ or a middle finger movement with a ‘1’).

Each trial began with the presentation of a fixation cross for 500 milliseconds (ms). A face image was presented at the top of the screen after the fixation cross for 500 ms, followed by the neutral hand image

Table 1
Primary and secondary questions of interest for experiments 1, 2, and 3 of this paper.

Experiment no.	Primary questions	Secondary questions
1	a) Is there a sex difference on the automatic imitation task when imitative and spatial effects are combined? b) Is there a sex difference on the flanker task?	a) Is automatic imitation invariant to stable traits of personality even when the context is more social? b) Are flanker and imitation task compatibility effects correlated with each other? c) Are there sex differences in how stable dimensions of personality predict automatic imitation?
2	a) Is there a sex difference on the automatic imitation task when the imitative component is orthogonal to the spatial component? b) Is there a sex difference on the flanker task?	a) Is automatic imitation invariant to individual differences in empathy and alexithymia? b) Are flanker and imitation task compatibility effects correlated with each other? c) Are there sex differences in how empathy and alexithymia predict automatic imitation?
3	a) Is there a sex difference and/or an in-group bias (based on sex) on the imitative compatibility effect? b) Is there a sex difference and/or an in-group bias (based on sex) on the spatial compatibility effect?	a) Is automatic imitation invariant to stable traits of personality when it is independent of spatial confounds? b) Are there sex differences in how stable dimensions of personality predict automatic imitation?

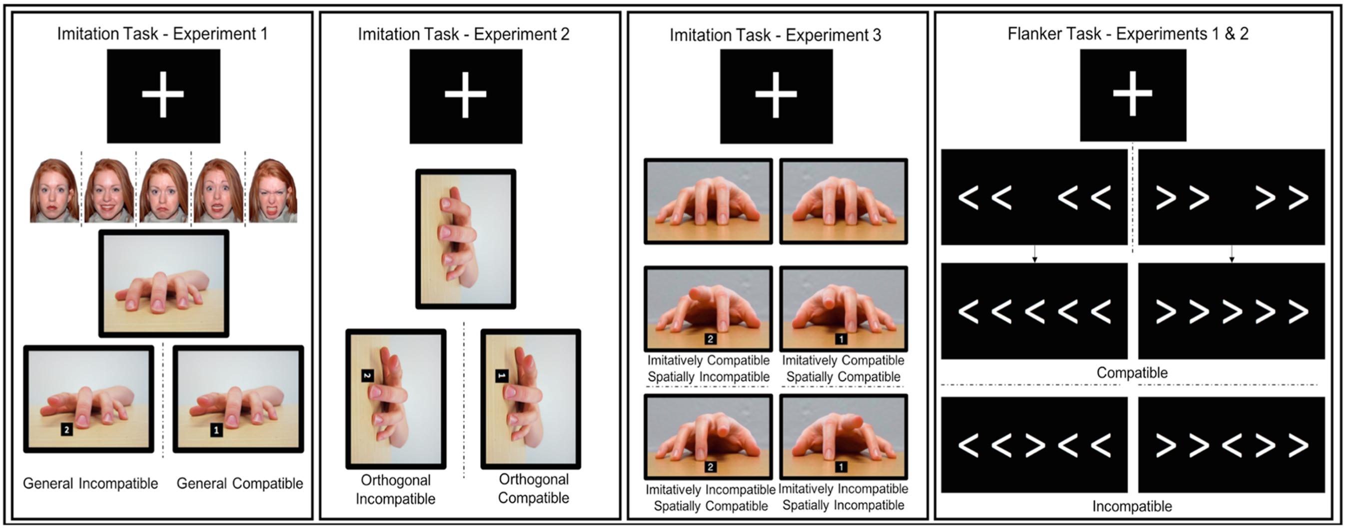


Fig. 1. Imitation and flanker tasks. Stimuli and trial design for the imitation and flanker tasks in Experiments 1, 2, and 3. Flanker tasks were similar for Experiment 1 and 2. In Experiment 1, in the imitation task, hand stimuli were preceded by a face depicting either a neutral, happy, sad, fearful, or angry image. In Experiment 2, hand stimuli were presented orthogonal to the participant's response hand, and in Experiment 3, both left- and right-hand images were used in order to measure imitative and spatial effects independent of each other.

in the centre of the screen. The face image remained on the screen above the neutral hand and target hand image for the remainder of the trial. The neutral hand was presented for a random inter-stimulus interval (ISI) of 500, 700, or 1000 ms, followed by the target hand image. The succession of neutral and target hand images was such that it produced apparent motion of either an index or middle finger lift simultaneously with the presentation of the number cue. The target hand image remained on the screen until the participant made a response (but no longer than 2000 ms). The total trial length varied depending on the ISI, but was never longer than 3500 ms. Trials were pseudo-randomised in such a way that no more than four identical trials were presented consecutively. There were four blocks of 50 trials each which included 25 compatible trials and 25 incompatible trials with equal number of trials per face image.

2.2.2.2. *Flanker task*

The flanker task was based on the paradigm developed by Eriksen and Eriksen (1974; Fig. 1). The stimuli consisted of five equally sized and spaced white arrows on a black background. Participants were instructed to respond to the direction of the central arrow – they were asked to press key 'm' with their right index finger if the central arrow pointed to the right, and press key 'n' with their left index finger if the central arrow pointed to the left. The direction of the flanker arrows was either compatible (< < < < < OR > > > > >) or incompatible (< < > < < OR > > < > >) to the central arrow direction. This produced four trial types and two conditions (compatible and incompatible).

Each trial started with a fixation cross for 800 ms, 1000 ms, or 1200 ms. The flanker arrows then appeared on the screen for 100 ms, followed by the central arrow in between the flankers. The five arrows remained on the screen until the participant responded (but for no longer than 1600 ms). Participants were first presented with the fixation cross for 800 ms, 1000 ms, or 1200 ms, followed by the presentation of the four flanker arrows for 100 ms. Total trial length was never longer than 2900 ms. Trials were pseudo-randomised in such a way that no more than four identical trials were presented consecutively. Each participant did one block of 64 trials, with 32 compatible, and 32 incompatible trials. Further, in this experiment, we addressed an additional unrelated question – in half of the compatible and incompatible trials, flanker arrows flipped arrow direction during the trial between their initial presentation on the screen and the appearance of the central arrow. However, as we were interested in the basic compatibility effect, we collapsed trials across conditions irrespective of whether they changed direction mid trial or not.

Participants first completed the automatic imitation task, followed by the flanker task. Before starting each task, they completed a 10-trial practice block.

2.2.2.3. *Questionnaires*

Participants also completed a range of self-report questionnaires which included the Mini International Personality Item Pool (mini IPIP; Donnellan, Oswald, Baird, & Lucas, 2006); the Short Autism Spectrum Quotient (AQ-10 Adult; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Allison, Auyeung, & Baron-Cohen, 2012), the Brief Schizotypal Personality Questionnaire (SPQ-B; Raine & Benishay, 1995), the Narcissistic Personality Inventory (NPI-16; Ames, Rose, & Anderson, 2006), the Hypersensitivity Narcissism Scale (HSNS; Hendin & Cheek, 1997), the Interpersonal Reactivity Index (IRI; Davis, 1980), and the Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994). For more details on the measures used and how the questionnaires were scored, see the Supplementary material.

2.2.3. *Data analysis*

Accuracy on the imitation task was recorded as the proportion of trials that were correct i.e. when participants lifted the correct finger in response to the number cue. Reaction time (RT) was recorded as time taken from target onset to participant's response. Only correct trials were used to calculate RT. Trials on which participants responded incorrectly, i.e. lifted the wrong finger, responded after 2000 ms, or before target onset were all excluded from the analysis (5.64%).

Accuracy on the flanker task was recorded as the proportion of trials that were correct i.e. when participants pressed the correct button in response to the central arrow direction. RT was calculated as the time taken from target onset (i.e. presentation of the arrow) to when the participant made a response. Only correct trials were used to calculate RT. Trials on which participants responded incorrectly, i.e. lifted the wrong finger, responded after 1600 ms, or before target onset were all excluded from the analysis (4.77%). Compatibility effects were calculated for both the flanker and imitation tasks by subtracting reaction times on compatible trials from reaction times on incompatible trials.

Data was analysed as follows: first, for both the RT and accuracy data on the imitation task, a 2 (compatibility: incompatible, compatible) × 5 (emotion: neutral, sad, happy, fearful, angry) repeated measures ANOVA was performed to investigate whether facial cues signalling emotional states modulated the compatibility effect on the imitation task. Second, on both RT and accuracy data, for the flanker and imitation tasks separately, a 2 (compatibility: compatible, incompatible) × 2 (sex: male, female) mixed ANOVA was performed in order to investigate whether the compatibility effect on the imitation and flanker tasks varies as a function of sex. For all ANOVAs, we report Greenhouse Geisser corrected values if the assumption of sphericity is violated. Third, in order to investigate whether the flanker and imitation compatibility effects were correlated, a one-tailed Pearson's correlation was performed. A positive correlation would suggest that the two compatibility effects were related to each other.

Based on prior research (Brass et al., 2000; Eriksen & Eriksen, 1974; Heyes, 2011), we expected a main effect of compatibility on the 2 × 2 ANOVA such that RT would be higher, and accuracy would be lower on incompatible trials compared to compatible trials. In support of our hypothesis, we also expected a Compatibility ∗ Sex interaction such that the compatibility effect would be higher for females as compared to males. The interaction effect was central to testing our primary hypothesis, and thus, we calculated compatibility effects for male and female participants separately by computing the mean difference and 95% confidence intervals between compatible and incompatible conditions. In order to directly estimate the size of the difference in compatibility effects between males and females, we then again computed the mean difference and 95% confidence interval. We used one-tailed 95% confidence intervals as we had a directional hypothesis that females would have a higher compatibility effect than males on both the imitation and flanker tasks (Butler et al., 2015; Stoet, 2010; Clayson et al., 2011). Further, we plot the group-average data and individual participant data in the Results section to ensure that any effects that go in a direction contrary to our hypothesis can be visualised. We also provide all raw data online so that researchers can test for alternative hypotheses.

We also report standardised effect sizes for ANOVA using partial eta-squared (η_p^2), for independent samples *t*-tests using Cohen's *d* and for paired samples t-tests using Cohen's d_z (Cohen, 1992; Lakens, 2013). We also report and interpret the point and interval estimate using 95% CIs for effect sizes of interest in line with recent suggestions (Amrhein, Greenland, & McShane, 2019; Cumming, 2012). In order to quantify the evidence for a null hypothesis over the experimental hypothesis (where a null result was found using null hypothesis significance testing; NHST), we calculated the Bayes factor (BF_{01}) by performing a Bayesian independent samples *t*-test to investigate the sex difference between the sexes. We used default priors in JASP for the independent samples t-test, that is, a Cauchy distribution3 with spread r set to 1/√2. The Bayes factor was interpreted using benchmark criteria from Jeffreys (1961). Bayesian analyses, Cohen's d and d_z, as well as 95% CIs were calculated using JASP (JASP Team, 2018).

Further, as previous research demonstrated that the compatibility effect (as measured on the SRC imitation task) is invariant to stable traits of personality (Butler et al., 2015), we also investigated whether personality variables influenced automatic imitation by using multiple regression analyses. We introduced a more social context to the task by introducing facial cues signalling emotional expressions simultaneously with the hand images. Based on prior work, we predicted that facial cues signalling positive emotions would increase automatic imitation compared to neutral and negative emotional expressions (Rauchbauer et al., 2015; Butler et al., 2016). Following Butler et al. (2015), participant sex was coded as −1 for males and +1 for females. Raw scores on all questionnaires and mean reaction time were centred i.e. the group mean of a variable was subtracted from each individual score on that variable. We set up a base model comprising mean RT (collapsed across all conditions), participant sex, and the mean RT ∗ sex interaction, as these factors have been shown to explain variance in automatic imitation previously (Butler et al., 2015). We then individually tested the contribution of each of the personality measures by adding them to the base model in separate hierarchical multiple regression analyses. By doing so, we are able to address the extent to which personality measures predict variance in the SRC imitation task above and beyond the base model. To transparently visualise and report the data, we also include zero-order correlations between personality measures and performance on the SRC imitation task. As sex differences have been previously found on personality measures (Schmitt, Realo, Voracek, & Allik, 2008), we computed sex ∗ trait interaction terms for all personality variables, and evaluated them in separate multiple regression models.

For all questionnaires used, we report reliability information as indexed by Cronbach's alpha (Cronbach, 1957). For the imitation and flanker task compatibility effects, we report the internal consistency of the measure by calculating split-half reliability using a permutation-based split-half approach with 5000 random splits (Parsons, Kruijt, & Fox, 2019) using the splithalf package in R. All reliability analyses are reported in the Supplementary Material.

Throughout the paper, following Gigerenzer (2018), we avoid interpreting results based solely on *p* values and a binary distinction between “significant” and “non-significant”. Instead, we base the direction and strength of our interpretation on a range of metrics, which include a p value and an associated measure of sensitivity (power), effect sizes in original and standardised units along with a measure of precision using 95% confidence intervals (Cumming, 2012). Further, we run multiple replication and extension experiments (Zwaan et al., 2018) and we meta-analyse the main effects across experiments (Cumming, 2012), both of which help to further aid the confidence that we can have in our findings and the credibility of the conclusions more generally.

2.3. Results

2.3.1. Automatic imitation task

2.3.1.1. Accuracy. Average accuracy on the imitation task was above 90% for both males and females on both compatible and incompatible conditions (Supplementary Fig. 1, Supplementary Table 1). A 2 (compatibility: compatible, incompatible) × 5 (emotion: neutral, sad, happy, fearful, angry) ANOVA showed no main effect of emotion (F (4,696) = 0.50, p = 0.729, η_p^2 = 0.003) and no Compatibility ∗ Emotion interaction (F (4,656) = 1.20, p = 0.31, η_p^2 = 0.007). Thus, for all further analyses of accuracy, trials are collapsed across all emotion conditions.

The 2 (compatibility: compatible, incompatible) × 2 (sex: male, female) mixed ANOVA showed a main effect of compatibility such that participants were more accurate on compatible trials than incompatible trials (F(1,173) = 258.09, $p < 0.001$, η_p^2 = 0.60; Supplementary Fig. 1). The effect size for the main effect of compatibility is conventionally considered to be large. For the main effect of sex (F (1,173) = 0.22, p = 0.64, η_p^2 = 0.001) and the Compatibility ∗ Sex interaction (F(1,173) = 0.60, p = 0.44, η_p^2 = 0.003), the effect sizes were close to zero (Supplementary Table 2).

2.3.1.2. Reaction time. Average reaction times on the imitation task for both males and females on both compatible and incompatible conditions were between 485 and 585 milliseconds (Fig. 2, Supplementary Table 1). A 2 (compatibility: compatible, incompatible) × 5 (emotion: neutral, sad, happy, fearful, angry) ANOVA showed no main effect of emotion (F (4,696) = 1.81, p = 0.127, η_p^2 = 0.004). Importantly, the effect size for the Compatibility ∗ Emotion interaction was close to zero (F (4,696) = 0.40, p = 0.796, η_p^2 = 0.002, see Supplementary Fig. 2). Thus, for all further analyses of RT, trials are collapsed across all emotion conditions.

The 2 (compatibility: compatible, incompatible) × 2 (sex: male, female) mixed ANOVA showed a main effect of compatibility such that participants were slower to respond on incompatible trials than

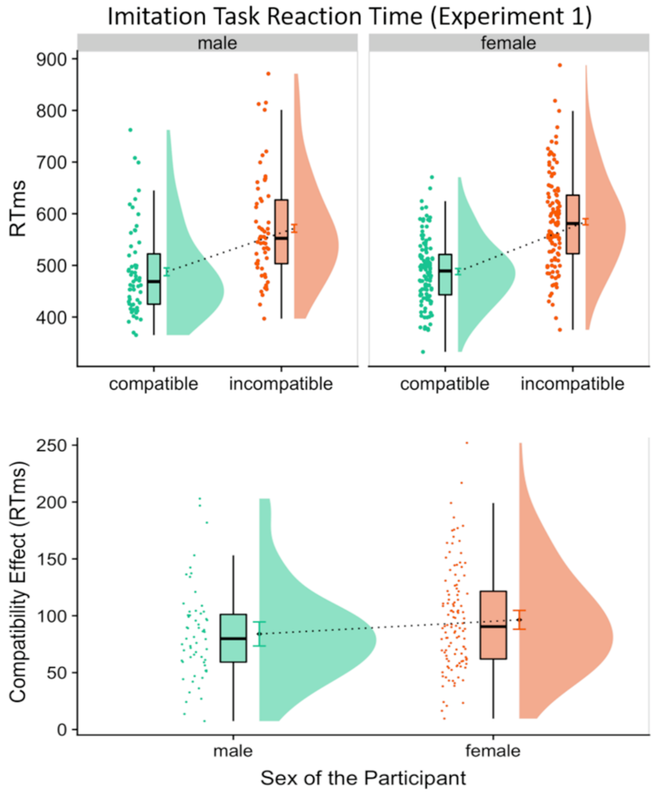


Fig. 2. Experiment 1 – Raincloud plots (Allen et al., 2018) showing imitation task reaction time. Reaction time is reported in milliseconds (ms). The upper panel shows mean reaction times for compatible and incompatible conditions for both males and females. The lower panel shows the compatibility effect for both males and females. The compatibility effect is calculated by subtracting reaction times on compatible trials from incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms = reaction time in milliseconds.

compatible trials (F(1,173) = 669.77, $p < 0.001$, $\eta_p^2 = 0.80$; Fig. 2). The effect size for the main effect of compatibility is conventionally considered to be large. The effect size for the main effect of sex was close to zero with a p-value of $p = 0.61$ (F(1,173) = 0.26, p = 0.61, $\eta_p^2 = 0.001$). The Compatibility ∗ Sex interaction had a small effect size (F (1,173) = 3.16, $p = 0.08$, $\eta_p^2 = 0.018$; Supplementary Table 2).

To further explore our primary research question regarding sex differences in the imitation task, compatibility effects were computed separately for males and females, and then compared to each other. For both males and females, compatibility effects had a large standardised effect size (Cohen's $d_z > 2.07$) with the lower bound of the 95% confidence interval at 1.68 or higher. When compatibility effects for males and females were directly compared to each other, we found a mean difference in the direction that was predicted (females > males). Indeed, the compatibility effect for females was 12.40 ms higher than males and the lower bound of the 95% confidence interval was 0.87 ms (Mean Difference = 12.40 ms, 95% CI[0.87, ∞], Cohen's d = 0.28; Fig. 2, Table 2A). The standardised effect of d = 0.28 is conventionally considered a small-to-medium effect, and the lower bound of the 95% confidence interval for the effect size was just above zero (95% CI [0.02, ∞]). Thus, these findings suggest that performance on the imitation task differs as a function of sex in a manner that is consistent with our predictions, such that females had a greater compatibility effect than males.

2.3.2. Flanker task

2.3.2.1. Accuracy. Average accuracy on the flanker task was above 94% for both males and females on both compatible and incompatible conditions (Supplementary Fig. 3, Supplementary Table 1). A 2 (compatibility: compatible, incompatible) × 2 (sex: male, female) mixed ANOVA was performed. The effect size for the main effect of compatibility was close to zero (F(1,180) = 2.24, $p = 0.136$, $\eta_p^2 = 0.01$). Effect sizes for the main effect of sex and (F(1,180) = 0.04, $p = 0.85$, $\eta_p^2 \leq 0.001$) and the Compatibility ∗ Sex interaction (F (1,180) = 0.09, $p = 0.759$, $\eta_p^2 < 0.001$) were also close to zero with p-values > 0.1 (Supplementary Table 2).

Table 2
Compatibility effects for the imitation and flanker tasks across Experiments 1, 2, and 3.

A) Experiment 1	Compatibility effect for the imitation and flanker task			
	Mean difference (ms)	95% CI	Cohen's d_z/d	BF_{01}
Imitation task (general compatibility effect)				
Males	84.03	(75.19, ∞)	2.07 [1.68, ∞]	
Females	96.43	(89.49, ∞)	2.14 [1.86, ∞]	
Females − males	12.40	(0.87, ∞)	0.28 [0.02, ∞]	
Flanker task (flanker compatibility effect)				
Males	51.17	(42.46, ∞)	1.28 [0.98, ∞]	
Females	56.79	(51.42, ∞)	1.58 [1.36, ∞]	
Females − males	5.62	(−4.14, ∞)	0.15 [−0.11, ∞]	3.85

B) Experiment 2	Compatibility effect for the imitation and flanker task			
	Mean difference (ms)	95% CI	Cohen's d_z/d	BF_{01}
Imitation task (orthogonal compatibility effect)				
Males	25.79	(21.72, ∞)	1.02 [0.82, ∞]	
Females	32.77	(28.82, ∞)	1.28 [1.07, ∞]	
Females − males	6.98	(1.34, ∞)	0.27 [0.05, ∞]	
Flanker task (flanker compatibility effect)				
Males	93.88	88.89	3.11 [2.71, ∞]	
Females	94.87	89.96	2.98 [2.62, ∞]	
Females − males	0.98	(−6.01, ∞)	0.03 [−0.19, ∞]	6.58

C) Experiment 3	Spatial and imitative compatibility effects (collapsed across all levels of stimulus sex)			
	Mean difference (ms)	95% CI	Cohen's d_z/d	BF_{01}
Spatial compatibility (spatially incompatible − spatially compatible)				
Males	30.74	(27.64, ∞)	1.18 [1.02, ∞]	
Females	37.28	(33.97, ∞)	1.37 [1.20, ∞]	
Females − males	6.53	(2.02, ∞)	0.24 [0.07, ∞]	
Imitative compatibility (imitatively incompatible − imitatively compatible)				
Males	7.02	(4.50, ∞)	0.33 [0.21. ∞]	
Females	8.35	(5.95, ∞)	0.42 [0.30, ∞]	
Females − males	1.33	(−2.15, ∞)	0.06 [−0.10, ∞]	4.95

N.B. Compatibility effects for males and females, as well as the difference between males and females, for the imitation and flanker tasks are reported for Experiments 1, 2, and 3, along with 95% CIs, effect sizes and BF_{01}. Abbreviations: ms = milliseconds, CI = confidence intervals, BF = Bayes Factor.

2.3.2.2. Reaction time. Mean reaction time on the flanker task for both males and females on both compatible and incompatible conditions was between 420 and 495 milliseconds (Fig. 3, Supplementary Table 1). A 2 (compatibility: compatible, incompatible) × 2 (sex: male, female) mixed ANOVA (Fig. 3) showed a main effect of compatibility such that participants were slower to respond on incompatible trials than

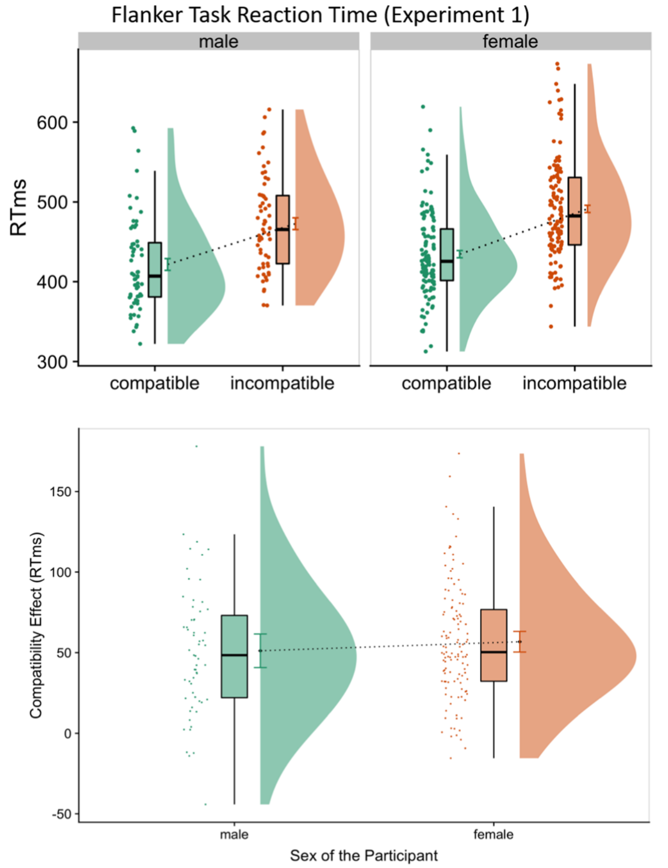


Fig. 3. Experiment 1 – flanker task reaction time. Reaction time is reported in milliseconds (ms). The upper panel shows mean reaction times for compatible and incompatible conditions for both males and females. The lower panel shows the compatibility effect for both males and females. The compatibility effect is calculated by subtracting reaction times on compatible trials from incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms = reaction time in milliseconds.

compatible trials (F(1,180) = 334.15, $p < 0.001$, $\eta_p^2 = 0.65$). The effect size for the main effect of compatibility is conventionally considered to be large. The main effect of sex had a small effect size and showed that overall females were slower than males on the flanker task (F(1,180) = 3.40, $p = 0.08$, $\eta_p^2 = 0.02$). The Compatibility ∗ Sex interaction had an effect size close to zero with a *p*-value of 0.34 (F (1,180) = 0.90, $p = 0.34$, $\eta_p^2 = 0.005$; see Supplementary Table 2).

To further compare with the automatic imitation task, compatibility effects in the flanker task were computed separately for males and females, and then compared to each other. For both males and females, compatibility effects had a large effect size (Cohen's $d_z > 1.2$) with the lower bound of the 95% confidence interval at 0.98 or higher. When compatibility effects for males and females were directly compared to each other, there was a trend toward females showing a higher compatibility effect than males by 5.62 ms, with the lower bound of the 95% confidence interval at −4.14 ms below zero (Mean Difference = 5.62 ms, 95% CI[−4.14, ∞], Cohen's d = 0.15, 95% CI [−0.11]; see Fig. 3, Table 2A). The effect size was small, with the lower bound of the 95% confidence interval at −0.11. Thus, a reasonable estimate for the mean difference between males and females on the flanker compatibility effect ranges from −4.14 ms to 5.62 ms, with one possibility being a small difference between females and males, such that females may show a higher compatibility effect than males. However, a Bayesian independent samples *t*-test showed that the data was 3 to 4 times more likely under the null hypothesis than the alternative hypothesis ($BF_{01} = 3.85$).

2.3.3. Correlational analysis

In order to investigate whether the flanker and imitation compatibility effects were correlated, a one-tailed skipped correlation was performed. To do so, only those participants who performed both the tasks were included in the analysis (*N* = 165). The skipped correlation analyses were performed using a Matlab-based toolbox (Mathworks Inc., MA; http://sourceforge.net/projects/robustcorrtool/, Pernet, Wilcox, & Rousselet, 2013). Skipped correlation takes into consideration the overall structure of the data, and protects against bivariate outliers. In order to perform a skipped correlation analysis, we first tested the assumptions of normality and homogeneity of variance. The Henze-Zirkler's multivariate normality test (Trujillo-Ortiz, Hernandez-Walls, Barba-Rojo, & Cupul-Magana, 2007) indicated that the data was close to normally distributed, and the test for heterogeneity indicated that the data have the same variance. Next, we estimated the robust centre of the data using the minimum covariance determinant (MCD) estimator. The MCD estimator is considered to be a robust estimator of the scatter and location of multivariate data (Rousseeuw, 1984; Rousseeuw & Van Driessen, 1999; Verboten & Hubert, 2005). Bivariate outliers were then identified by using a projection technique – data points were orthogonally projected by lines joining each data point to the robust centre of the data cloud. Five bivariate outliers were removed using the box-plot rule relying on the interquartile range (Carling, 2000), and skipped correlation was computed on the remaining data. Following guidelines put forward previously (Pernet et al., 2013), as the data were close to being normally distributed, we used a skipped Spearman correlation analysis. A one-tailed skipped Spearman correlation analysis showed a small positive correlation between the imitation and flanker compatibility effects, which did not pass our statistical threshold, with the lower bound of the 95% CI extended below zero (r (160) = 0.04, 95% CI [−0.12, ∞]; see Fig. 4).

Our results thus suggest that the compatibility effects across these tasks were largely unrelated, and participants with greater interference in one task did not experience a greater interference in the other task.

2.3.4. Multiple regression analyses

We also investigated the relationship between stable personality measures and the general compatibility effect as measured on the SRC task. We tested for multicollinearity between our variables by calculating the variance inflation factor (VIF). VIF is the ratio of overall model variance compared to the variance of a model that includes only a single predictor variable. A high VIF for a predictor variable indicates that the variable is highly collinear with other variables that are included in the model. VIF was calculated for every single predictor variable separately using the function lm() and vif() in R. In the current analyses, VIF for each predictor variable was < 2 indicating a very low level of multicollinearity.. The base model included mean RT, sex, and a meanRT ∗ sex interaction term. The base model explained 33.6% of the variance in the congruency effect (F(3,171) = 28.88, $p < 0.001$, $R^2 = 0.34$, $f^2 = 0.51$) and indicated a medium effect size (Cohen, 1992). Mean RT predicted the compatibility effect, with increasing CE as mean RT increases (B = 0.27, SEB = 0.03, t(171) = 8.02, $p < 0.001$, 95% CI [0.20; 0.34]). In addition to mean RT, sex marginally predicted the compatibility effect (B = 5.40, SEB = 2.88, t (171) = 1.87, $p = 0.06$, [−0.29; 11.10]) with a higher compatibility effect for females than males. The mean RT ∗ sex interaction was also a predictor (B = 0.10, SEB = 0.03, t(171) = 3.01, $p = 0.003$, [0.03; 0.17]), suggesting that increases in mean RT predicted larger increases in the compatibility effect for females (B = 0.37, SEB = 0.04, t (171) = 8.43, $p = 0.001$) compared to males (B = 0.17, SEB = 0.05, t (171) = 3.32, $p = 0.001$). Results from the base model are very similar to the results of prior work using a same SRC task and analytical approach (Butler et al., 2015).

Agreeableness, extraversion, grandiose and vulnerability narcissism, empathy, autistic-like and schizotypal traits, and alexithymia did not predict the general compatibility effect above and beyond the base

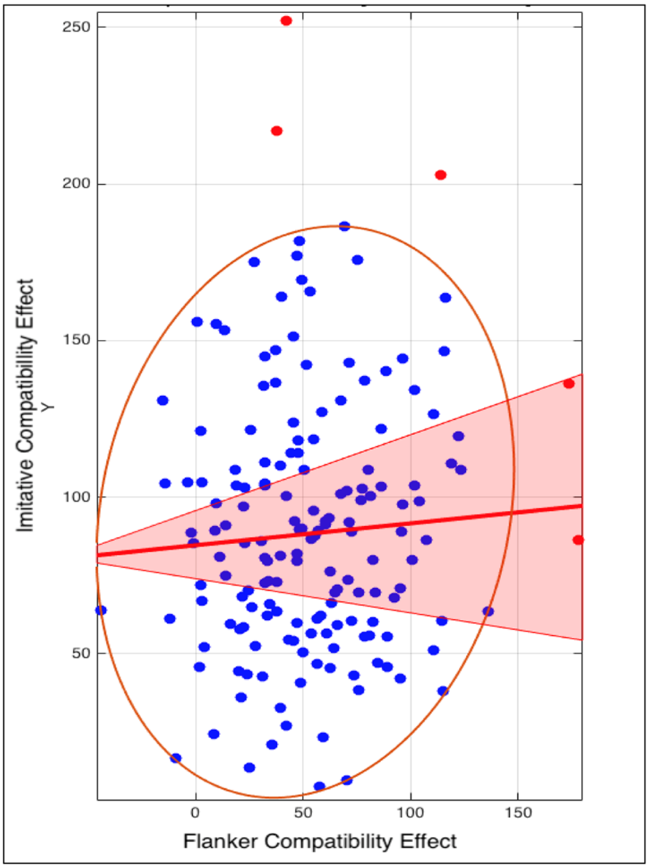


Fig. 4. Experiment 1 - correlation analysis. A skipped Spearman correlation showed a small positive correlation between the flanker and imitative compatibility effects that does not pass our statistical thresholding. Abbreviations: RTms (reaction time in milliseconds). Dots in red are the bivariate outliers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

model (all p's $>$ 0.11, all CIs overlapping with zero; see Fig. 5). Effect sizes attributable to the addition of the personality variables (beyond the base model) indicated very small effects (Cohen's f^2 for all models < 0.01; Cohen, 1992). The multiple regression models are summarized in Supplementary Table 3. Zero-order correlations are also consistent with the findings from the multiple regression analyses, such that there are no relationships between stable personality measures and CE (see Supplementary Table 4, Supplementary Fig. 4).

To evaluate the sex ∗ trait interaction terms, we computed additional models – each model consisted of the base model, one trait predictor, and the sex ∗ trait interaction term. For alexithymia, when the sex ∗ trait term was included in the model, the model explained 34.9% of the variance. The sex ∗ alexithymia interaction term marginally predicted the compatibility effect (B = 0.48, SEB = 0.27, t(169) = −1.71, p = 0.09, 95% CI [−0.07, 1.03]), and explained an additional 1.3% of the variance (ΔR2 = 0.013, F(1,169) = 2.93, p < 0.09). A decrease in alexithymia marginally predicted an increase in the compatibility effect only for males (B = −0.80, SEB = 0.47, t(169) = −1.72, p = 0.09) and not for females (B = 0.15, SEB = 0.30, t(169) = 0.49, p = 0.62; see Fig. 6A). The effect size attributable to the addition of alexithymia and the sex ∗ trait interaction term was very small (Cohen's f^2 = 0.02).

For empathy, when the sex ∗ trait term was included in the model, the model explained 36.9% of the variance. Sex ∗ empathic concern predicted the compatibility effect above and beyond the base model (B = 1.62, SEB = 0.74, t(167) = 2.17, p = 0.03, 95% CI [0.15, 3.10]) and explained an additional 1.8% of the variance (ΔR2 = 0.018, F(1,167) = 4.73, p = 0.031). Sex ∗ perspective taking (B = −1.91, SEB = 0.81, t(167) = −2.34, p = 0.02, 95% CI [−3.52, −0.30]) predicted the compatibility effect above and beyond the base model and explained an additional 2.1% of the variance (ΔR2 = 0.021, F(1,167) = 5.49, p = 0.021). An increase in empathic concern marginally predicted a decrease in the compatibility effect for males (B = −2.24, SEB = 1.23, t(167) = −1.82, p = 0.07) whereas in females, there was a trend for an increase in empathic concern predicting an increase in the compatibility effect (B = 1.01, SEB = 0.85, t(169) = 1.19, p = 0.24). An increase in perspective taking predicted a decrease in the compatibility effect in females (B = −1.90, SEB = 0.85, t(167) = −2.24, p = 0.026). In males, there was a trend for an increase in perspective taking predicting an increase in the compatibility effect (B = 1.92, SEB = 1.39, t(167) = 1.38, p = 0.169; see Fig. 6A). The effect size attributable to the addition of empathy and the sex ∗ trait interaction term was small (Cohen's f^2 = 0.05). None of the other sex ∗ trait interaction terms predicted the compatibility effect above and beyond the base model (Cohen's f^2 for all models < 0.02, Supplementary Fig. 5, Supplementary Table 5).

2.4. Discussion

The results demonstrate a sex difference in the general compatibility effect on the imitation task such that females showed a higher general compatibility effect than males, thus replicating the direction of results found previously (Butler et al., 2015; Genschow et al., 2017). A similar sex difference, however, was not found on the flanker task. Moreover, flanker and general compatibility effects were largely unrelated to each other. At first glance, therefore, this suggests that the sex difference may be tied to a form of cognitive control that is not shared between the two tasks, such as social (imitative) control.

Before we can make firm conclusions regarding the type of cognitive structure supporting the sex difference, however, we first consider some limitations of these results. First, the general compatibility effect is a sum of both spatial and imitative features. Participants respond with their right hand to a number cue – they are asked to lift their index or middle finger (for ‘1’ or ‘2’ respectively) while simultaneously observing a left hand making either the same or different finger movements. However, in this task, the observed and executed movements are not just imitatively compatible or incompatible, but also on the same or different side of space i.e. spatially compatible or incompatible. Thus, the task measures a general compatibility effect i.e. it does not measure the control of automatic imitation or the imitative compatibility effect independent of spatial compatibility effects (Catmur & Heyes, 2011). Thus, the sex difference may reflect a difference in spatial compatibility with respect to a finger location in space, as opposed to specifically in imitation control.

A second limitation to these initial conclusions is that the flanker task used in the current experiment employed fewer trials than those used in previous studies where a sex difference was found (e.g. Stoet, 2010; Clayson et al., 2011). Therefore, a lack of sex difference might reflect a lack of precision in measuring the effect. This might also explain why we did not find a main effect of compatibility on the accuracy data in the flanker task. Thus, although the current experiment employed a larger sample size than previous studies using the flanker task, we are still cautious to interpret the lack of evidence for the sex difference in the first experiment.

We also addressed an additional unrelated question in the current experiment. In half of the compatible and incompatible trials, flanker arrows flipped arrow direction during the trial between their initial presentation on the screen and the appearance of the central arrow. However, when we analysed the flip and no-flip trials separately, findings were similar and the direction of the results did not change. Therefore, we do not think that the flipping of the arrow direction had any consequences on the measure of interest.

Further, in the current experiment, we did not find any effect of the

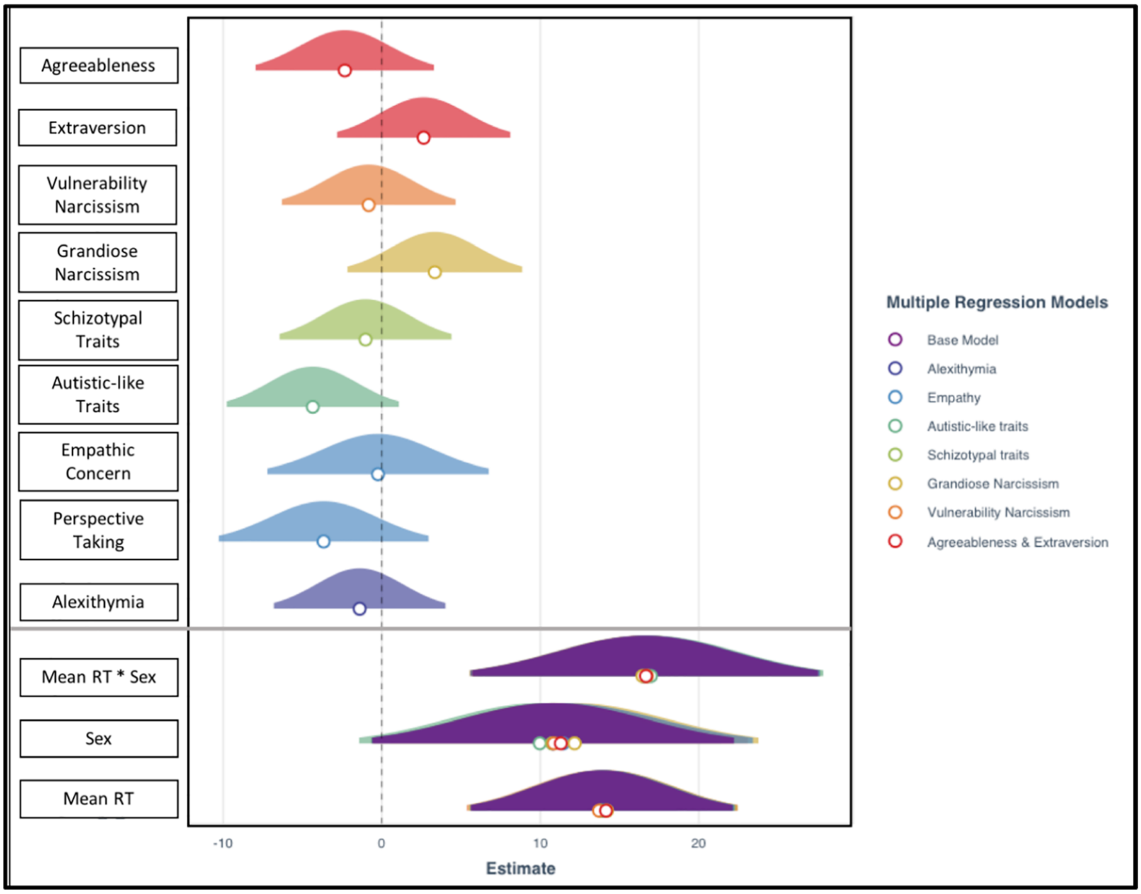


Fig. 5. Experiment 1 – multiple regression analyses. For the outcome variable of general compatibility, values of standardised coefficients are plotted for each predictor variable (personality trait) along with their corresponding uncertainties (95% confidence interval width for a normal distribution for each estimate). Coefficients are standardised by dividing by two standard deviation units according to Gelman, Jakulin, Pittau and Su (2008). The base model consists in the bottom three predictor variables (depicted in violet) – mean RT, Sex, and meanRT ∗ Sex. Abbreviations: RT = Reaction Time. *N.B.* The circles on the purple distributions represent the standardised co-efficients for the main effect of meanRT, main effect of sex, and meanRT ∗ sex interaction respectively for each of the models tested. That is why there are multiple circles for components of the base model because the base model was part of all the models tested (that is, one model for each personality trait). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

type of emotional expression on automatic imitation. These findings add to previous research that shows mixed evidence for a link between the emotional expression of the interacting partner and the tendency to automatically imitate (Grerucci et al., 2013; Crescentini et al., 2011; Rauchbauer et al., 2015). Finally, if we turn to consider the effects of stable personality measures, a clear picture begins to emerge. Even in a more socially meaningful context where emotional expressions are signalled, we further support the claim that imitative control in general (across the entire group of participants), shows a general invariance to stable dimensions of personality like narcissism, agreeableness, extraversion, autistic-like and schizotypal traits (Butler et al., 2015; Cracco et al., 2018). Of course, it is possible that the emotional expressions failed to add to the social context of the task in a meaningful way. The face image signalling the emotional expression was presented along with the hand image but they were two separate images. This ensured that participants also paid attention to the hand image which was the focus of the task. Further, it has been recently suggested that the link between inferring an emotional state and the corresponding facial movements or expressions is less clear than what has been previously suggested (Barrett, Adolphs, Marsella, Martinez, & Pollak, 2019). It is possible, therefore, that participants ignored the face image, did not think of the hand as connected to the face image, and did not infer the emotional state of the stimuli. Even if this were true, however, we add a further large dataset to the prior work (Butler et al., 2015; Cracco et al., 2018), which all show that personality variables have little relationship to performance on the imitation task in general. We suggest that studies purporting alternative patterns of relationship between imitation and personality measures in general across the population (Chartrand & Bargh, 1999; Hogeveen & Obhi, 2013; Obhi et al., 2013) perform powerful replications to enable a cumulative science to develop (Munafò et al., 2017; Zwaan et al., 2018).

Although there were no clear main effects of personality across the entire group, there was some suggestive evidence that the effect of personality on imitation differed by sex. Given prior evidence linking automatic imitation and alexithymia, we expected that an increase in automatic imitation would be predicted by a decrease in alexithymia (Sowden et al., 2016). In the current experiment, this was true only for males, and not for females. We further predicted that an increase in empathic concern and perspective taking would predict an increase in automatic imitation (Chartrand & Bargh, 1999). However, the current findings suggest that a decrease in empathic concern predicts the compatibility effect in males, but not females, and a decrease in perspective taking predicts the compatibility effect in females, but not males. It has been suggested that males score higher on measures of alexithymia as compared to females (Levant, Hall, Williams, & Hasan, 2009), and females score higher on empathic concern and perspective taking as compared to males (Christov-Moore et al., 2014; Van der Graaff et al., 2014). We had no a priori hypotheses however as to whether females and males would show a link between personality and imitation in different directions. In addition, these sex ∗ trait

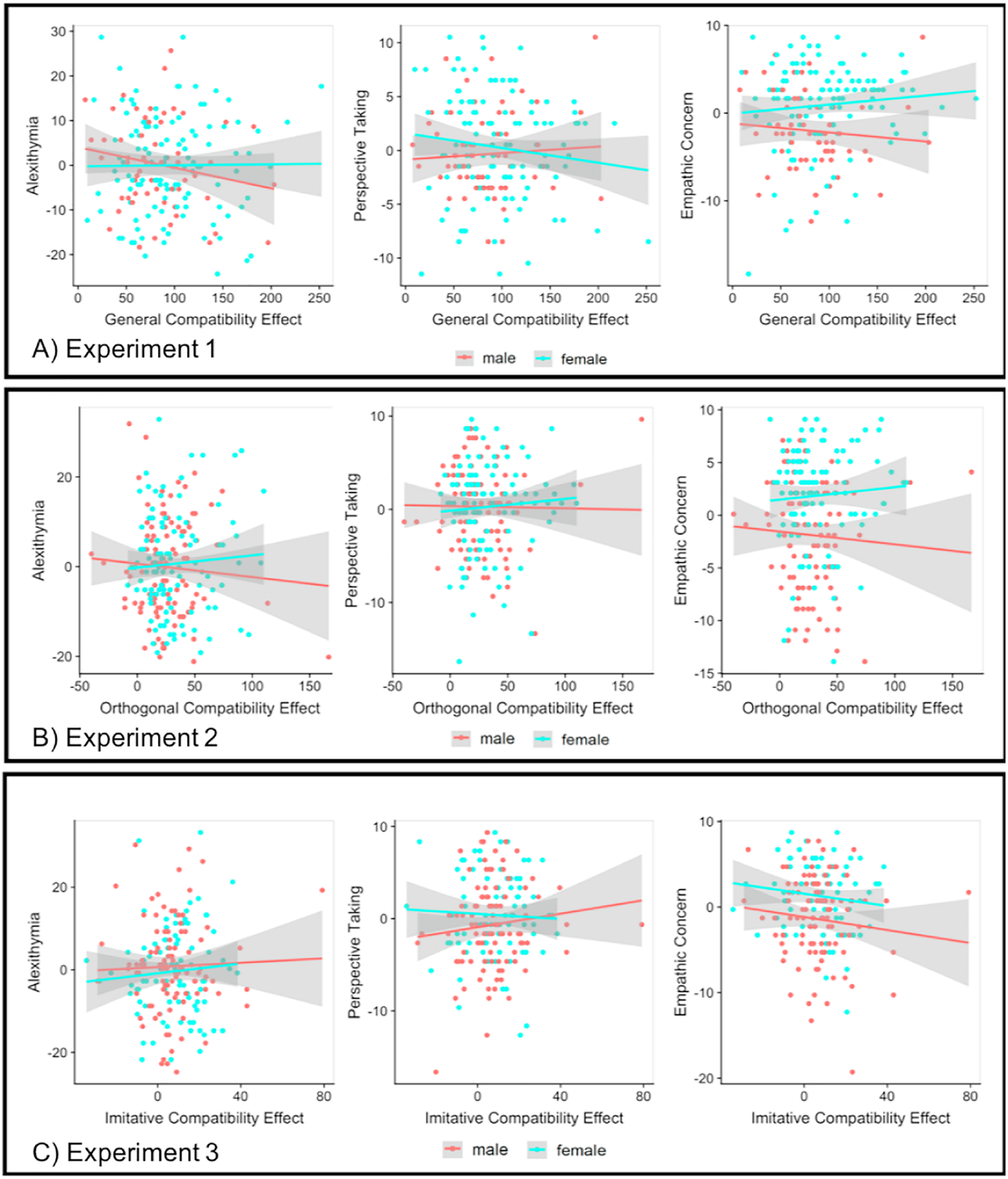


Fig. 6. Sex by trait interactions for Experiments 1, 2, 3. Sex by trait interactions for alexithymia, perspective taking, and empathic concern for Experiments 1 (A), 2 (B), and 3 (C). X axis denotes the imitative compatibility effect in milliseconds, and Y axis denotes mean centred scores on the personality traits.

interactions were small effects and contributed to only an additional 1.3% (alexithymia) and 3.3% (empathy) of the model. Thus, before making any firm conclusions, these results require replication in order to confirm that they do not reflect false positives as a result of sampling error.

A final potential limitation is that there were unequal samples in the first experiment – 59 males and more than a hundred females. However, we do not think that this affected overall findings because 1) although sample sizes are unequal, we do not violate the homogeneity of variance assumption for the ANOVA, 2) unequal sample sizes are problematic for factorial ANOVAs where there are unequal samples for two (or more) between-group variables (which is not the case in the current experiment; Kao & Green, 2008), and 3) we analyse our data using different approaches and find similar results from all analyses, thus lending further support to our findings. Even so, for all following experiments, we have kept sample sizes roughly equal.

Overall, therefore, these initial results from Experiment 1 demonstrate that cognitive control systems may operate differently across individuals on some (sex), but not other (personality), stable dimensions of individuals.

3. Experiment 2

3.1. Introduction

In the second experiment, we extend findings from Experiment 1 and address its limitations by making the following changes. First, in the automatic imitation task, stimuli were displayed orthogonal to the response hand in order to minimise the effect of spatial compatibility. Thus, instead of the general compatibility effect, we now investigate the sex difference on the orthogonal compatibility effect. The orthogonal compatibility effect allows us the measure automatic imitation dissociated from right-left spatial compatibility effects, thus allowing for a more precise measure of the imitative effect. Second, we again compare between males and females on the flanker task but increase the number of trials such that both the imitation and flanker tasks are equal. Similar to Experiment 1, we performed a correlational analysis to see whether flanker and orthogonal compatibility effects were related to each other or not.

In Experiment 1, three sex ∗ trait interactions, which covered empathic concern, perspective taking, and alexithymia, predicted the general compatibility effect. Thus, in order to further confirm these findings, we included empathy (empathic concern and perspective taking) and alexithymia measures in Experiment 2 to investigate whether these traits modulated the orthogonal compatibility effect.

3.2. Method

3.2.1. Participants

Two hundred and thirty-eight participants took part in this experiment for monetary compensation (£6) or course credit. All participants provided informed consent, had normal or corrected-to-normal vision, and were right-handed. Approval was obtained from the Research Ethics and Governance Committee of the School of Psychology at Bangor University. One participant was excluded because data on only half the trials was recorded on the flanker task.

Participants were excluded if performance was three standard deviations away from the group mean average performance per condition in terms of accuracy or reaction time ($N = 15$ for the imitation task, $N = 21$ for the flanker task). For the imitation task, the final sample included 223 participants (107 males, $\text{Mean}_{\text{age}} = 20$, $\text{SD}_{\text{age}} = 4.33$; Mean_{age} and SD_{age} are based on 203 participants as some participants did not enter their age in the demographic questionnaire). For the flanker task, the final sample included 217 participants (101 males, $\text{Mean}_{\text{age}} = 20.7$, $\text{SD}_{\text{age}} = 4.31$; Mean_{age} and SD_{age} are based on 198 participants).

3.2.2. Stimuli, tasks, and procedure

3.2.2.1. Automatic imitation task. The automatic imitation task was similar to the one used in Experiment 1, with the following changes: one, no face image was presented during the task (Fig. 1). Two, the hand stimuli were presented orthogonal to the response (Fig. 1). Three, there were 360 trials in total, which comprised six blocks of 60 trials, each of which included 30 compatible and 30 incompatible trials.

3.2.2.2. Flanker task. The flanker task was the same as Experiment 1 with only one change – participants completed 360 trials in total, with 6 blocks of 60 trials each (30 compatible and 30 incompatible trials; Fig. 1).

The order of the tasks was counterbalanced such that half the participants did the flanker task first, whereas the remaining half did the imitation task first.

3.2.2.3. Questionnaires. Participants also completed two self-report questionnaires which included the Interpersonal Reactivity Index (IRI; Davis, 1980), and the Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994). For more details on the measures used, see the Supplementary material.

3.2.3. Data analysis

Accuracy and RT on the imitation and flanker tasks were recorded in the same way as Experiment 1 and only correct trials were used to calculate RT. Trials on which participants responded incorrectly, i.e. lifted the wrong finger, responded after 2000 ms, or before target onset (imitation = 5.59%; flanker = 5.97%) were all excluded from the analysis.

Data were analysed in the same way as Experiment 1. For the imitation task, a Sex ∗ Compatibility interaction showing a higher compatibility effect for females compared to males would indicate that the sex difference on the imitation task persists even when stimuli are presented orthogonally to the response. Alternatively, similarly sized compatibility effects between the sexes would suggest that reducing the spatial component of the task largely removes the sex difference.

3.3. Results

3.3.1. Automatic imitation task

3.3.1.1. Accuracy. Average accuracy for both males and females for both compatible and incompatible trials was over 92% (see Supplementary Fig. 7, Supplementary Table 6). The 2 (compatibility: compatible, incompatible) × 2 (sex: male, female) mixed ANOVA showed a main effect of compatibility such that participants were more accurate on compatible trials than incompatible trials ($F(1,221) = 96.22$, $p < 0.001$, $\eta_p^2 = 0.30$). The effect size for the main effect of compatibility is conventionally considered to be large. The effect sizes for the main effect of sex ($F(1,221) = 1.87$, $p = 0.17$, $\eta_p^2 = 0.008$) and the Compatibility ∗ Sex interaction ($F(1,221) = 0.14$, $p = 0.71$, $\eta_p^2 < 0.001$) were close to zero with p-values > 0.1 (see Supplementary Table 7).

3.3.1.2. Reaction time. Mean reaction times were between 435 and 485 milliseconds for both males and females on both compatible and incompatible trials (see Fig. 7, Supplementary Table 6). The 2 (compatibility: compatible, incompatible) × 2 (sex: male, female) mixed ANOVA (Fig. 7) showed a main effect of compatibility such that participants were slower to respond on incompatible trials than compatible trials ($F(1,221) = 293.18$, $p < 0.001$, $\eta_p^2 = 0.56$). The effect size for the main effect of compatibility is conventionally considered to be large. The main effect of sex had a relatively small effect size and showed that females were generally slower than males ($F(1,221) = 4.23$, $p = 0.040$, $\eta_p^2 = 0.02$). There was a Compatibility ∗ Sex interaction and the effect size is conventionally considered to be a small effect ($F(1,221) = 4.17$, $p = 0.042$, $\eta_p^2 = 0.02$; Supplementary Table 7).

In order to interrogate our primary hypothesis regarding sex differences in the imitation task, we computed compatibility effects separately for males and females, and then compared them to each other. For both males and females, compatibility effects had a large effect size (Cohen's $d_z > 1.0$) and the lower bound of the 95% confidence interval was at least 0.82. When compatibility effects for males and females were directly compared to each other, we found a mean difference of 6.98 ms in the direction that was predicted i.e. the compatibility effect

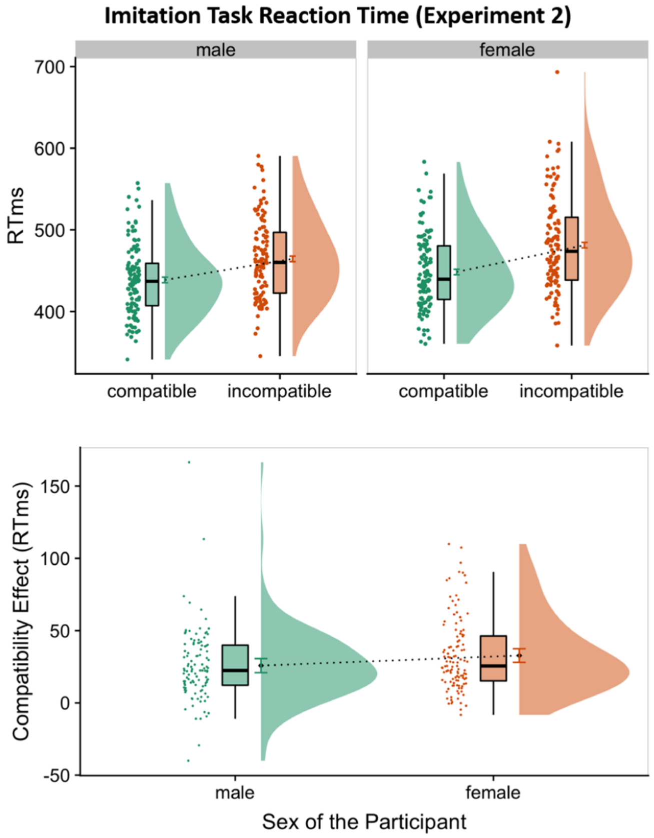


Fig. 7. Experiment 2 – imitation task reaction time. Reaction time is reported in milliseconds (ms). The upper panel shows mean reaction times for compatible and incompatible conditions for both males and females. The lower panel shows the compatibility effect for both males and females. The compatibility effect is calculated by subtracting reaction times on compatible trials from incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms = reaction time in milliseconds.

for females was greater than the compatibility effect for males with the lower bound of the 95% confidence interval above zero (Mean Difference = 6.98 ms, 95% CI[1.34, ∞], Cohen's d = 0.27 95% CI [0.05, ∞]; Fig. 7, Table 2B). The effect size was a small-to-medium effect, with the lower bound of the 95% CI at 0.05.

The absolute size of the difference between the sexes as measured in original units (i.e., ms) is smaller than Experiment 1, as the orthogonal compatibility effect is smaller than the general compatibility effect measured in Experiment 1. Indeed, when measured in original units, the compatibility effect in Experiment 2 is approximately half the size of Experiment 1 and the same is true for the sex difference in compatibility effect between the two experiments. However, the standardised effect size for the sex difference is nearly identical across the two experiments (Exp. 1 = 0.28; Exp. 2 = 0.27). Therefore, when measured in comparable units, which account for differences in absolute values, these results suggest that the sex difference measured is quite consistent across experiments. In sum, the orthogonal compatibility effect on the imitation task differed as a function of sex in the same manner and to a similar degree as Experiment 1, such that females had a greater orthogonal compatibility effect than males.

3.3.2. *Flanker task*

3.3.2.1. Accuracy. Average accuracy was over 88% for both males and females on both compatible and incompatible trials (see Supplementary Fig. 8, Supplementary Table 6). A 2 (compatibility: compatible, incompatible) × 2 (sex: male, female) mixed ANOVA showed a main effect of compatibility (F(1,215) = 151.33, $p < 0.001$, $\eta_p^2 = 0.41$).

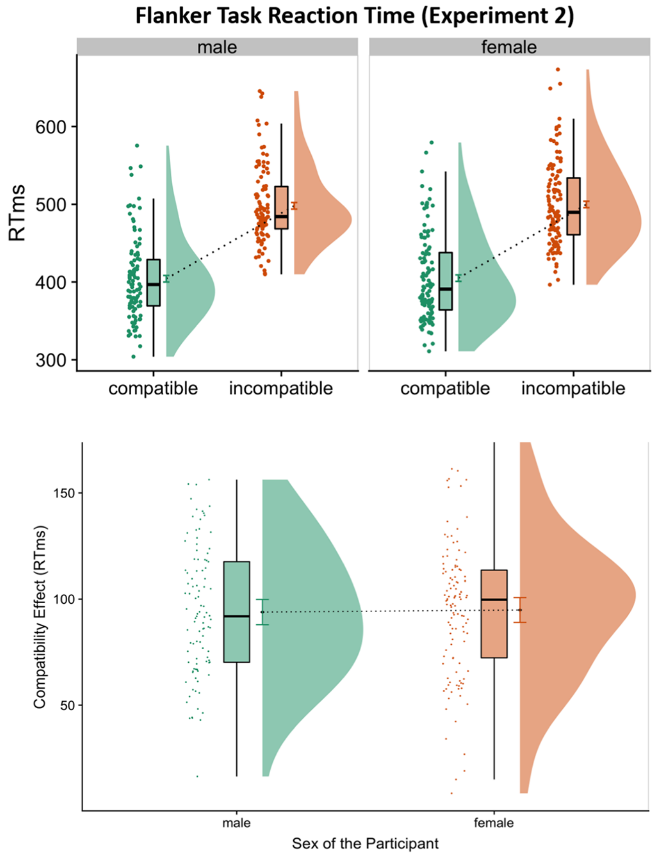


Fig. 8. Experiment 2 – flanker task reaction time. Reaction time is reported in milliseconds (ms). The upper panel shows mean reaction times for compatible and incompatible conditions for both males and females. The lower panel shows the compatibility effect for both males and females. The compatibility effect is calculated by subtracting reaction times on compatible trials from incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms = reaction time in milliseconds.

The main effect of compatibility showed that participants were more accurate on compatible trials compared to incompatible trials. The main effect of sex showed that females had lower accuracy overall compared to males (F(1,215) = 5.78, $p = 0.017$, $\eta_p^2 = 0.03$). The Compatibility ∗ Sex interaction (F(1,215) = 3.17, $p = 0.076$, $\eta_p^2 = 0.01$) showed that the difference in accuracy between compatible and incompatible trials was greater for females compared to males. The effect sizes for both the main effect of sex, and the interaction were relatively small (see Supplementary Table 7).

3.3.2.2. Reaction time. Mean reaction times for both males and females for both compatible and incompatible conditions was between 400 and 500 milliseconds (see Fig. 8, Supplementary Table 6). A 2 (compatibility: compatible, incompatible) × 2 (sex: male, female) mixed ANOVA (Fig. 8) showed a main effect of compatibility such that participants were slower to respond on incompatible trials than compatible trials (F(1,215) = 1986.89, $p < 0.001$, $\eta_p^2 = 0.90$). The effect size for the main effect of compatibility is conventionally considered to be large. The effect sizes for the main effect of sex (F(1,215) = 0.03, $p = 0.854$, $\eta_p^2 < 0.001$). and the Compatibility ∗ Sex interaction (F(1,215) = 0.05, $p = 0.816$, $\eta_p^2 < 0.001$) were close to zero with *p*-values > 0.8 (see Supplementary Table 7).

To explore sex differences in the flanker task further, compatibility effects were computed separately for males and females, and then compared to each other. For both males and females, compatibility effects had a large effect size (Cohen's $d_z > 2.9$) with the lower bound of the 95% confidence interval at least 2.62. When compatibility effects for males and females were directly compared to each other, females showed a higher compatibility effect than males, but the effect size was very small, with the lower bound of the 95% confidence interval reaching −6.01 ms (Mean Difference = 0.98 ms, 95% CI[−6.01, ∞], Cohen's d = 0.03, 95%CI[−0.19, ∞]; Fig. 8, Table 2B). The effect size was close to zero with the lower bound of the confidence interval at −0.19 (below zero). A Bayesian independent samples *t*-test showed that the null was 6 to 7 times more likely than the alternative hypothesis (BF_{01} = 6.58). Thus, although both males and females separately showed a compatibility effect, there was a negligible difference between males and females on the flanker compatibility effect.

3.3.3. Correlational analysis

In order to investigate whether the flanker and imitation compatibility effects were correlated, a one-tailed skipped correlation was performed. For the correlational analysis, only those participants who performed both the tasks were included in the analysis (*N* = 205). As in Experiment 1, we also performed a more robust correlation analysis. The data was not normally distributed, but was homoscedastic. Thus, we performed a skipped Spearman correlation analysis on 191 participants as 14 bivariate outliers were detected. Results indicated that flanker and imitation compatibility effects showed a weak positive correlation that did not pass our statistical threshold (Spearman r (191) = 0.07, 95% CI [−0.07, ∞]; Fig. 9). Our findings thus suggest that flanker and imitative compatibility effects are largely unrelated, and interference on one task did not predict interference on the other.

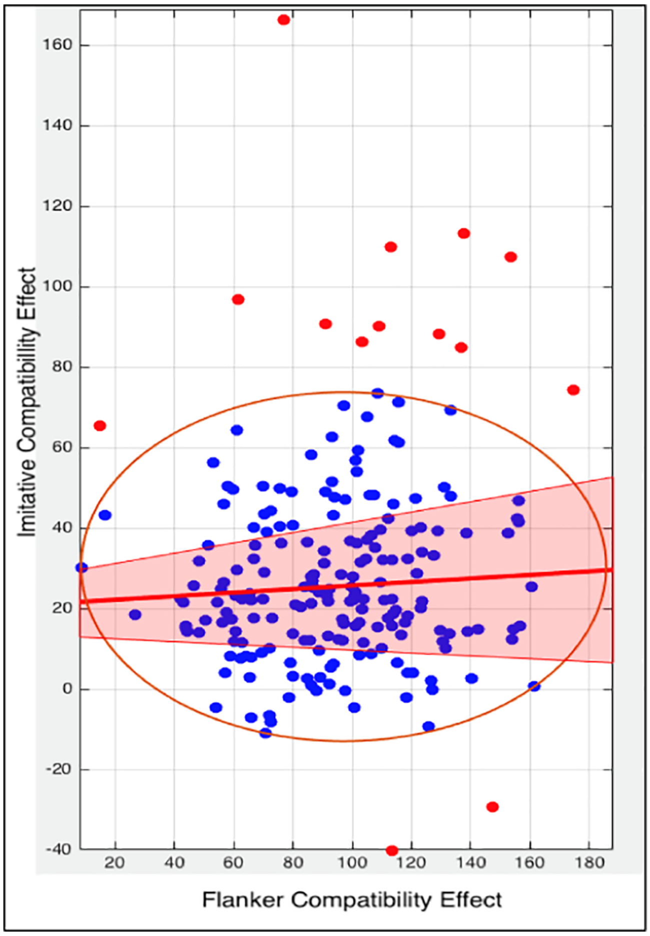


Fig. 9. Experiment 2 - correlation analysis. A skipped Spearman correlation showed a small positive correlation between the flanker and imitative compatibility effects that does not pass our statistical thresholding. Abbreviations: RTms (reaction time in milliseconds). Dots in red are the bivariate outliers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3.4. Multiple regression analyses

We also investigated the relationship between personality variables (empathy and alexithymia) and the orthogonal compatibility effect as measured on the SRC task. Tests for multicollinearity indicated that a very low level of multicollinearity was present (*VIF* for all predictor variables < 2). The base model (which included mean RT, sex, and the mean RT ∗ Sex interaction) explained 16.2% of the variance in the congruency effect (F(3,204) = 13.13, *p* < 0.001, R2 = 0.16, f^2 = 0.19) and indicated a medium effect. Mean RT predicted the orthogonal compatibility effect (B = 0.19, SEB = 0.03, t(204) = 5.36, p < 0.001, [0.12; 0.26]) with increasing CE as mean RT increased. Sex did not predict the orthogonal compatibility effect (B = 2.14, SEB = 1.67, t(204) = 1.27, *p* = 0.21, [−1.19; 5.47]). The mean RT ∗ sex interaction (B = 0.07, SEB = 0.03, t(204) = 1.90, *p* = 0.06, [−0.002; 0.13]) suggested that increases in mean RT predicted larger increases in the compatibility effect for females (B = 0.25, SEB = 0.05, t(204) = 5.45, p < 0.001) compared to males (B = 0.12, SEB = 0.05, t(204) = 2.33, p < 0.001).

Alexithymia and empathy (empathic concern and perspective taking) did not predict the orthogonal compatibility effect above and beyond the base model (all p's > 0.03, all CIs overlapping with zero; see Fig. 10). Effect sizes attributable to the addition of the personality variables (beyond the base model) indicated extremely small effects (Cohen's f^2 = 0.001 for alexithymia and Cohen's f^2 = 0.005 for empathy). The multiple regression models are summarized in Supplementary Table 8. Zero-order correlations are also consistent with the findings from the multiple regression analyses (see Supplementary Table 9, Supplementary Fig. 9).

To evaluate the sex ∗ trait interaction terms, we computed additional models – each model consisted of the base model, one trait predictor, and the sex ∗ trait interaction term. None of the sex ∗ trait interaction terms predicted the orthogonal compatibility effect above and beyond the base model (all ps > 0.3, all CIs overlapping with zero; Fig. 6B, Supplementary Fig. 10). Effect sizes attributable to the addition of the sex ∗ trait interaction terms (beyond the base model) indicated extremely small effects (Cohen's f^2 = 0.01 for both alexithymia and empathy). The multiple regression models are summarized in Supplementary Table 10.

3.4. Discussion

As in Experiment 1, results indicated a clear sex difference in the orthogonal compatibility effect such that females showed a higher orthogonal compatibility effect compared to males on the automatic imitation task. The sex difference persisted on the imitation task in spite of presenting stimuli orthogonal to the response. However, this sex difference on the RT compatibility effect was not found on the flanker task even after increasing the number of trials. Further, the correlational analysis suggested that flanker and orthogonal compatibility effects were only marginally correlated with each other and explained only 0.8% of the variance. Thus, greater interference on one task is able to predict only a very small amount of interference on the other task.

Thus, across Experiments 1 and 2, we show a lack of consistent evidence for a sex difference in the flanker task. However, the interpretation of the sex difference on the imitation task still has two potential limitations. One, while the presentation of orthogonal stimuli reduces spatial compatibility effects on the left-right axis, they do not rule out the possibility of orthogonal spatial compatibility effects i.e. the propensity of participants to show an advantage for an up-right and down-left pairing (Weeks & Proctor, 1990; Cho & Proctor, 2003; Weeks, Proctor, & Beyak, 1995). For instance, when stimuli were presented orthogonal to the response hand (see Fig. 1), the index finger was always below the middle finger, and the participant's index finger was to the left side of space. Thus, a preference for responding to “up” stimuli

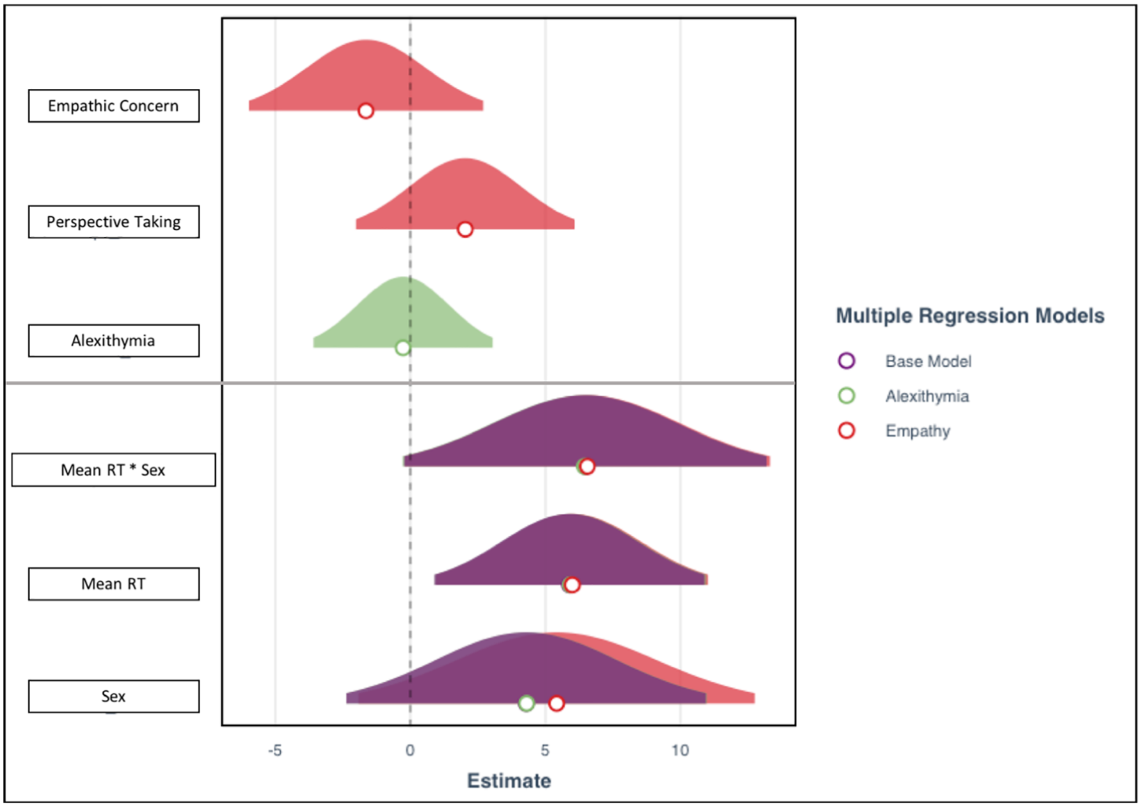


Fig. 10. Experiment 2 – multiple regression analyses. For the outcome variable of orthogonal compatibility, values of standardised coefficients are plotted for each predictor variable (personality trait) along with their corresponding uncertainties (95% confidence interval width for a normal distribution for each estimate). Coefficients are standardised by dividing by two standard deviation units according to Gelman (2008). The base model consists in the bottom three predictor variables (depicted in violet) – mean RT, Sex, and meanRT ∗ Sex. Abbreviations: RT = Reaction Time. *N.B.* The circles on the purple distributions represent the standardised co-efficients for the main effect of meanRT, main effect of sex, and meanRT ∗ sex interaction respectively for each of the models tested. That is why there are multiple circles for components of the base model because the base model was part of all the models tested (that is, one model for each personality trait). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with a right response and "down" stimuli with a left response may be observed along with imitative effects on the automatic imitation task used in the current experiment. Thus, the sex difference may reflect a difference on orthogonal spatial effects as opposed to purely imitative effects.

Two, the stimuli used in both Experiments 1 and 2 were those of a female model. We did not manipulate the sex of the stimulus, and therefore, the sex difference can either reflect a genuine difference between males and females, or an in-group bias. A difference between male and female participants (irrespective of the sex of the stimulus) would reflect distinct (or partially distinct) cognitive mechanisms underlying imitative or spatial control as a function of sex. On the contrary, an in-group bias or own-sex bias would suggest that sex differences as evidenced previously on the automatic imitation task (Butler et al., 2015; Genschow et al., 2017) do not actually reflect a sex difference - females show a higher compatibility effect because they favour members of the in-group i.e. of their own sex compared to members of the out-group i.e. of the opposite sex (Brown, 1995; Gleibs et al., 2016; Rauchbauer et al., 2015; Rudman & Goodwin, 2004).

Finally, in terms of personality measures, empathy and alexithymia (and sex ∗ trait interactions) did not modulate the orthogonal compatibility effect. Although we found suggestive evidence in Experiment 1 for a small link between personality traits (alexithymia and empathy) and imitation that differed between the sexes, the current experiment did not replicate these findings. Therefore, overall, these results provide limited support for a link between personality traits and automatic imitation, and confirm and replicate findings from previous large sample studies (Butler et al., 2015; Cracco et al., 2018) that suggest automatic imitation is largely invariant to stable traits of personality.

4. Experiment 3

4.1. Introduction

Experiment 3 addressed two remaining issues. First, we measured the imitative compatibility effect independently from the spatial compatibility effect, in order to estimate whether the sex difference reflects a spatial or more specialised (social) aspects of cognitive control. Second, we assessed the extent to which the sex difference reflects a basic difference between males and females and/or an in-group bias based on sex.

To separate imitative and spatial components of the task, we used a modified version of the SRC task of automatic imitation that allowed us to manipulate imitative and spatial effects separately (Bertenthal, Longo, & Kosobud, 2006; Boyer, Longo, & Bertenthal, 2012; Catmur & Heyes, 2011). A sex difference on spatial compatibility alone would indicate that the sex difference observed in Experiments 1 and 2 can be explained by differences associated with processing spatial information. Alternatively, a sex difference on imitative compatibility alone, would suggest that greater compatibility effects for females reflects a difference in the control of automatic imitation specifically.

To compare a sex difference account with an in-group bias account of our findings so far, we manipulated the sex of the stimuli used in the SRC task and again tested male and female participants. A greater compatibility effect for females for female stimuli compared to male stimuli would indicate that an own-sex bias contributes to the sex difference observed on the automatic imitation task. Alternatively, a sex difference on the task and relative invariance to the sex of the stimuli would suggest that there is a basic control mechanism that differs between males and females that seems resistant to possible contextual factors, such as group biases.

In order to investigate whether personality variables influence automatic imitation, in Experiment 3, we included all personality variables included in Experiment 1 (alexithymia, empathy, autistic-like and schizotypal traits, narcissism, extraversion, and agreeableness). In Experiment 1 and 2, the compatibility effect measured on the imitation task was a composite of spatial and imitative effects. Therefore, the invariance of the compatibility effect may be related to spatial effects as opposed to imitative effects. Therefore, we included all the personality measures in order to investigate whether imitative compatibility when measured independently of spatial effects is also invariant to stable personality traits.

4.2. Method

4.2.1. Participants

Two hundred and one participants took part in this experiment for monetary compensation (£6) or course credit. All participants provided informed consent, had normal or corrected-to-normal vision, and were right-handed. Approval was obtained from the Research Ethics and Governance Committee of the School of Psychology at Bangor University. Participants were excluded if performance was 3 standard deviations away from the group mean average performance per condition in terms of accuracy or reaction time on the imitation task ($N = 12$). The final sample included 189 participants (97 males, $\mathrm{Mean}_{\mathrm{age}} = 21.4$, $\mathrm{SD}_{\mathrm{age}} = 4.08$, age range = 18 to 42) ($\mathrm{Mean}_{\mathrm{age}}$ and $\mathrm{SD}_{\mathrm{age}}$ are based on 182 participants as 7 participants did not enter their age in the demographics questionnaire).

4.2.2. Stimuli, tasks, and procedure

4.2.2.1. Automatic imitation task. The automatic imitation task was similar to the one used in Experiment 2, with the following changes: one, stimuli were not presented orthogonally to the response. Two, we calculated an imitative compatibility effect independent of the spatial compatibility effect (Catmur & Heyes, 2011). For this, both left- and right-hand images were used as stimuli, but participants always responded with their right hand. This resulted in eight trial types and four conditions of interest (Fig. 1):

1. imitatively and spatially compatible (for example, when participants are cued to lift their index finger, and watch an index finger lift of the left hand, the observed finger movement is both spatially and imitatively compatible to the executed movement),
2. imitatively and spatially incompatible (for example, when participants are cued to lift their index finger, and watch a middle finger lift of the left hand, the observed finger movement is both spatially and imitatively incompatible to the executed movement),
3. imitatively compatible and spatially incompatible (for example, when participants are cued to lift their middle finger, and watch a middle finger lift of the right hand, the observed finger movement is imitatively compatible, but spatially incompatible to the executed movement),
4. imitatively incompatible and spatially compatible (for example, when participants are cued to lift their middle finger, and watch an index finger lift of the right hand, the observed finger movement is imitatively incompatible, but spatially compatible to the executed movement).

Thus, participants performed the same (imitatively compatible) or different (imitatively incompatible) movement on the same (spatially compatible) or different (spatially incompatible) side of space.

A third change in comparison to Experiment 2, is that in order to investigate whether the sex difference was due to an own-sex bias, the hand stimuli presented included 4 female and 4 male hands. The hand stimuli were chosen based on a pilot study. In the pilot study (see Supplementary material), eighteen hand stimuli were rated by 51 participants on a scale of 1 to 9 with one being most masculine, 5 being neutral, and 9 being most feminine. Four hand stimuli rated as most masculine, and four hand stimuli rated as most feminine were chosen for the current experiment. There were 360 total trials, with 90 trials per condition. Timing information and pseudo-randomisation was the same as in Experiment 1 and 2.

4.2.2.2. Questionnaires. Participants also completed a range of self-report questionnaires which included the Mini International Personality Item Pool (mini IPIP; Donnellan et al., 2006); the Short Autism Spectrum Quotient (AQ-10 Adult; Baron-Cohen et al., 2001), the Brief Schizotypal Personality Questionnaire (SPQ-B; Raine & Benishay, 1995), the Narcissistic Personality Inventory (NPI-16; Ames et al., 2006), the Hypersensitivity Narcissism Scale (HSNS; Hendin & Cheek, 1997), the Interpersonal Reactivity Index (IRI; Davis, 1980), and the Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994). For more details on the measures used, see the Supplementary material.

In order to confirm that participants perceived male and female stimuli differently, participants also rated the hand stimuli used in the experiment after they completed the task. Participants were asked to rate the stimuli on a scale of 1 to 9, with 1 being extremely masculine and 9 being extremely feminine.

4.2.3. Data analysis

Accuracy and RT on the imitation task were recorded in the same way as Experiment 1 and 2 and only correct trials were used to calculate RT. Trials on which participants responded incorrectly, i.e. lifted the wrong finger, responded after 2000 ms, or before target onset (7.41%) were all excluded from the analysis.

A key aim of our study was to investigate whether the sex difference and/or in-group bias exists in imitative and/or spatial compatibility effects (and not whether/how such differences differ between the two types of effect). For this purpose, therefore, we performed analyses separately on the spatial and imitative compatibility effects. For each compatibility effect separately, we performed a 2 (compatibility: incompatible, compatible) × 2 (stimulus sex: male hand, female hand) × 2 (participant sex: male, female) mixed ANOVA on the RT and accuracy data. Based on prior research (Catmur & Heyes, 2011; Darda et al., 2018; Gowen, Bolton, & Poliakoff, 2016; Marsh, Bird, & Catmur, 2016), we expected a main effect of spatial and imitative compatibility such that RT would be higher, and accuracy would be lower on spatially incompatible trials compared to spatially compatible trials, and on imitatively incompatible trials compared to imitatively compatible trials.

In addition, a Sex ∗ Compatibility interaction for spatial compatibility (such that females show a higher spatial compatibility effect that males) would be expected if the sex difference observed in Experiments 1 and 2 was largely driven by the spatial component of the task. In contrast, a Sex ∗ Compatibility interaction for imitative compatibility (such that females show a higher imitative compatibility effect than males) would suggest that the sex difference is largely a reflection of the imitative component of the task.

Alternatively, if the sex difference in the spatial or imitative compatibility effect is because of an own-sex bias, we would expect a three-way interaction (Sex ∗ Compatibility ∗ Stimulus Sex) such that females would be more interfered by a female stimulus, and males would be more interfered by a male stimulus i.e. females would show a higher compatibility effect than males for female stimuli compared to male stimuli.

As in Experiment 1, the interaction effect was central to testing our primary hypotheses, and thus, we calculated compatibility effects for male and female hand stimuli separately and independently for both male and female participants. To do so, we computed the mean difference and 95% confidence intervals between compatible and incompatible conditions across the levels of stimulus sex and participant sex. Spatial compatibility was calculated as RT on spatially incompatible trials minus RT on spatially compatible trials. Imitative compatibility was calculated as RT on imitatively incompatible trials minus imitatively compatible trials. In order to directly estimate the size of the difference in spatial and imitative compatibility effects between males and females, we then again computed the mean differences between the sexes and 95% confidence intervals. We used one-tailed 95% confidence intervals as we had a directional hypothesis that females would have a higher spatial or imitative compatibility effect than males.

For the secondary analyses, multiple regression analyses were performed in the same way as Experiments 1 and 2 in order to investigate whether personality variables and sub-clinical traits modulate automatic imitation when measured independent of spatial effects.

4.3. Results

4.3.1. Spatial compatibility

4.3.1.1. Accuracy. Average accuracy was over 92% for both males and females for all conditions of compatibility and stimulus sex (see Supplementary Fig. 11, Supplementary Table 11). A 2 (compatibility: incompatible, compatible) × 2 (stimulus sex: male hand, female hand) × 2 (participant sex: male, female) mixed ANOVA showed a main effect of compatibility such that participants were more accurate on compatible trials than incompatible trials ($F(1,187) = 563.35$, $p < 0.001$, $\eta_p^2 = 0.75$; Supplementary Fig. 11). The effect size of the main effect of compatibility was large. The main effect of stimulus sex suggested that participants were more accurate when observing male hand stimuli as compared to female hand stimuli ($F(1,187) = 335.47$, $p < 0.001$, $\eta_p^2 = 0.64$). The Compatibility ∗ Stimulus Sex interaction suggested that the difference in accuracy between incompatible and compatible trials was overall bigger for female stimuli compared to male stimuli (($F(1,187) = 202.31$, $p < 0.001$, $\eta_p^2 = 0.52$), Supplementary Fig. 11). All other main effects and interactions' effect sizes were relatively small with high *p*-values (see Supplementary Table 12).

4.3.1.2. Reaction time. Mean reaction times for both males and females on all conditions of compatibility and stimulus sex were between 415 and 475 milliseconds (see Fig. 11, Supplementary Table 11). A 2 (compatibility: incompatible, compatible) × 2 (stimulus sex: male hand, female hand) × 2 (participant sex: male, female) mixed ANOVA (Fig. 11) showed a main effect of compatibility such that participants were slower to respond on spatially incompatible trials compared to spatially compatible ($F(1,187) = 459.71$, $p < 0.001$, $\eta_p^2 = 0.71$). The effect size of the main effect of compatibility was large. The main effect of stimulus sex had a medium effect size, and suggested that overall participants responded slower to female hand stimuli than male hand stimuli ($F(1,187) = 5.63$, $p = 0.019$, $\eta_p^2 = 0.03$).

The Sex ∗ Compatibility interaction was a small-to-medium effect with a p-value of 0.04 ($F1,187) = 4.24$, $p = 0.041$, $\eta_p^2 = 0.02$). To interrogate the sex difference in spatial compatibility, we computed the difference in compatibility effects between males and females, collapsed across all conditions of stimulus sex. Females showed a higher compatibility effect than males by 6.53 ms, and the lower bound of the 95% CI was over zero at 2.02 ms (Mean difference = 6.53 ms, 95% CI [2.02, ∞], Cohen's $d = 0.24$, 95% CI [0.07, ∞]; Fig. 13A, Table 2C). The effect size was a small-to-medium effect, and the lower bound of the 95% CI was above zero at 0.07.

The three-way (Compatibility ∗ Stimulus Sex ∗ Sex) interaction ($F(1,187) = 1.77$, $p = 0.185$, $\eta_p^2 = 0.01$) showed a trend such that females had a higher compatibility effect for female hand stimuli compared to male hand stimuli, and males had a higher compatibility effect for male hand stimuli compared to female hand stimuli, although the effect size was close to zero (Fig. 11). All other effect sizes for main effects or interactions were close to zero (see Supplementary Table 12).

In order to investigate whether an in-group bias explains the sex difference in spatial compatibility, we computed compatibility effects on all levels of participant sex and stimulus sex. For both males and females, spatial compatibility effects were present when observing both male (Cohen's $d_z > 1.25$) as well as female stimuli (Cohen's $d_z > 1.10$). There was a trend for females showing a higher compatibility effect for female stimuli compared to male stimuli (Mean difference = 2.74 ms, 95% CI [−2.40, ∞], Cohen's $d_z = 0.09$, 95% CI [−0.08, ∞]), and for males showing a higher compatibility effect for male stimuli compared to female stimuli (Mean difference = 3.19, 95% CI [−2.12, ∞], Cohen's $d_z = 0.10$; 95% CI[−0.07, ∞]), but these were relatively small effect sizes (see Fig. 11).

4.3.2. Imitative compatibility

4.3.2.1. Accuracy. Average accuracy for both males and females for all conditions of stimulus sex and compatibility was above 87% (see Supplementary Fig. 12, Supplementary Table 11). A 2 (compatibility: incompatible, compatible) × 2 (stimulus sex: male hand, female hand) × 2 (participant sex: male, female) mixed ANOVA (Supplementary Fig. 12) showed a main effect of compatibility ($F(1,187) = 205.65$, $p < 0.001$, $\eta_p^2 = 0.52$) such that participants were more accurate on compatible trials than incompatible trials. The effect size of the main effect of compatibility was large. The main effect of stimulus sex ($F(1,187) = 335.47$, $p < 0.001$, $\eta_p^2 = 0.64$) suggested that participants were more accurate when observing male hand stimuli as compared to female hand stimuli. The Compatibility ∗ Stimulus Sex interaction ($F(1,187) = 162.98$, $p < 0.001$, $\eta_p^2 = 0.46$) suggested that the difference in accuracy between compatible and incompatible trials was bigger for female stimuli compared to male stimuli (see Supplementary Fig. 12). All other main effects and interactions were relatively small or close to zero with high *p*-values (see Supplementary Table 12).

4.3.2.2. Reaction time. Mean reaction times were between 430 and 460 milliseconds for both males and females on all conditions of compatibility and stimulus sex (see Fig. 12, Supplementary Table 11). A 2 (compatibility: incompatible, compatible) × 2 (stimulus sex: male hand, female hand) × 2 (participant sex: male, female) mixed ANOVA showed a main effect of compatibility ($F(1,187) = 54.96$, $p < 0.001$, $\eta_p^2 = 0.23$) such that participants were slower to respond on imitatively incompatible trials compared to imitatively compatible. The effect size of the main effect of compatibility was large. The main effect of stimulus sex was a small-to-medium effect with a p-value of 0.02 and suggested that participants responded slower to female hand stimuli than male hand stimuli ($F(1,187) = 5.70$, $p = 0.018$, $\eta_p^2 = 0.03$).

The effect size for the Sex ∗ Compatibility interaction was close to zero with a p-value of 0.52 ($F(1,187) = 0.41$, $p = 0.52$, $\eta_p^2 = 0.002$). Given the importance to our primary research question regarding sex differences in the compatibility effect, we interrogated the RT data further by computing the difference in compatibility effects between males and females, collapsed across all conditions of stimulus sex. Although females showed a marginally higher compatibility effect than males by 1.33 ms, the lower bound of the 95% CI was below zero at −2.15 ms. The effect size was small with the lower bound of the 95% CI below zero at −0.10 (Mean difference = 1.33, 95% CI [−2.15, ∞], Cohen's $d = 0.06$, 95% CI[−0.10, ∞]; see Fig. 13B, Table 2C).

The three-way (Compatibility ∗ Stimulus Sex ∗ Participant Sex) interaction ($F(1,187) = 3.86$, $p = 0.051$, $\eta_p^2 = 0.02$) was a relatively small effect. All other main effects or interactions had effect sizes close to zero with high p-values (see Fig. 12, Supplementary Table 12).

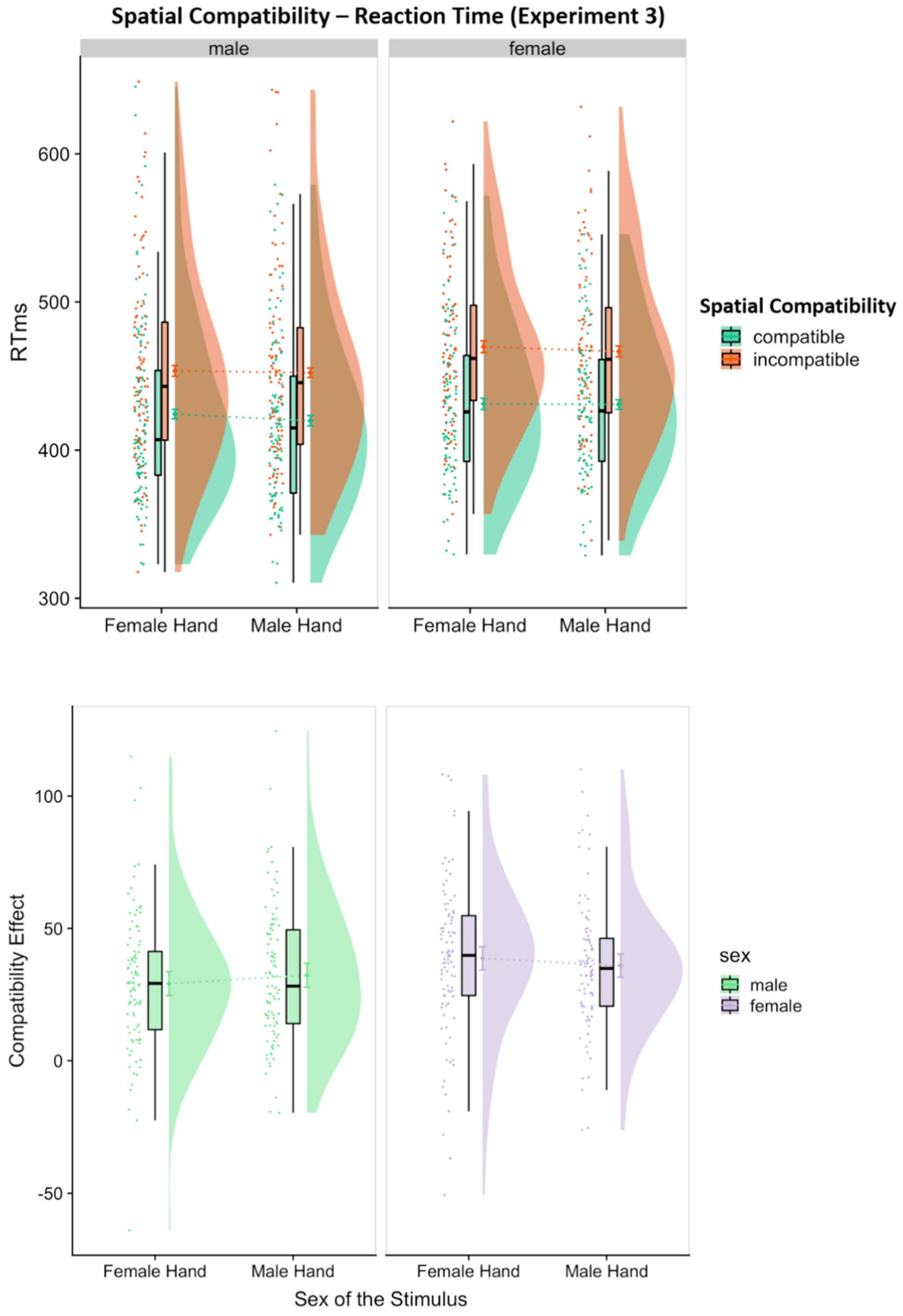


Fig. 11. Experiment 3 – spatial compatibility reaction time. Reaction time is reported in milliseconds (ms). The upper panel shows mean reaction times for compatible and incompatible conditions for both males and females, when responding to both male and female hand stimuli. The lower panel shows the compatibility effect for both males and females when responding to both male and female stimuli. The compatibility effect is calculated by subtracting reaction times on compatible trials from incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms = reaction time in milliseconds.

In order to investigate the three-way interaction and explore whether the sex difference can be explained by an in-group bias, we computed compatibility effects on all levels of participant sex and stimulus sex. For both males and females, imitative compatibility effects were present when observing both male (Cohen's $d_z > 0.2$) as well as female stimuli (Cohen's $d_z > 0.3$). However, there was not even a trend in the direction we predicted i.e. females did not show a higher compatibility effect for female stimuli compared to male stimuli (Cohen's $d_z = -0.10$), and for males showing a higher compatibility effect for male stimuli compared to female stimuli (Cohen's $d_z = -0.18$). On the contrary, the direction of the interaction was contrary to our hypothesis i.e. females showed a higher compatibility effect for male stimuli compared to female stimuli, and males showed a higher compatibility effect for female stimuli compared to male stimuli, but these effects were small (see Fig. 12). As such, not only are these effects relatively small, they are also inconsistent with the sex difference being a result of an ingroup bias based on the sex of the interaction partner.

In sum, our results indicated a sex difference in spatial compatibility, but not imitative compatibility. An in-group bias/own-sex bias did not explain the sex difference found in the spatial compatibility effect.

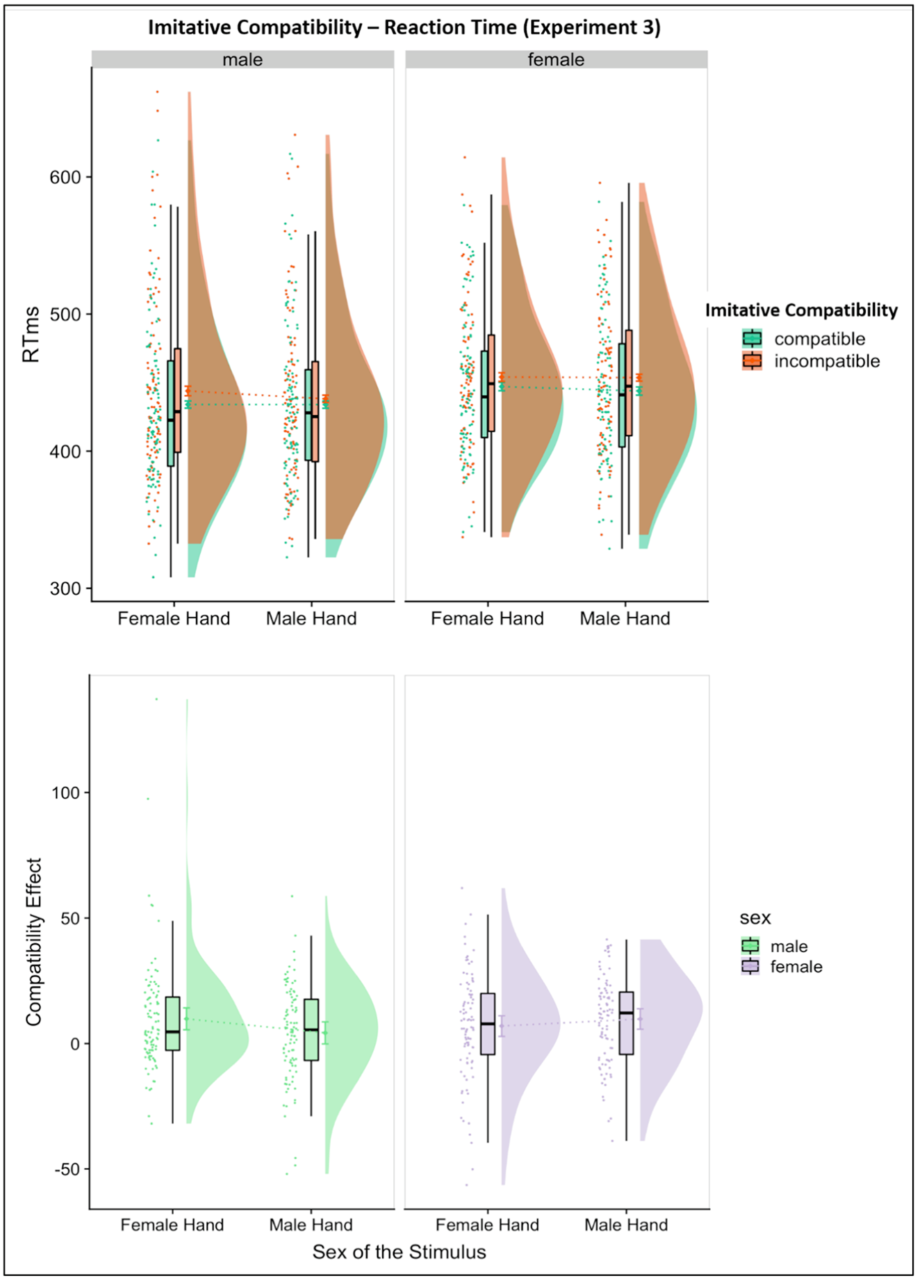


Fig. 12. Experiment 3 – imitative compatibility reaction time. Reaction time is reported in milliseconds (ms). The upper panel shows mean reaction times for compatible and incompatible conditions for both males and females when responding to both male and female hand stimuli. The lower panel shows the compatibility effect for both males and females when responding to both male and female hand stimuli. The compatibility effect is calculated by subtracting reaction times on compatible trials from incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms = reaction time in milliseconds.

4.3.3. *Multiple regression analyses*

We also investigated the relationship between stable personality measures and the imitative compatibility effect as measured on the SRC task independent of spatial effects. Tests for multicollinearity indicated that a very low level of multicollinearity was present (*VIF* for all predictor variables < 2). The base model (including sex, mean RT and the sex ∗ mean RT interaction) explained 4.59% of the variance in the imitative compatibility effect (F(3,181) = 2.90, *p* = 0.036, R2 = 0.04, f^2 = 0.05) indicating a small effect. Mean RT was a predictor (B = 0.05, SEB = 0.02, t(181) = 2.79, *p* = 0.006, [0.01; 0.09]), but both sex (B = 0.23, SEB = 1.04, t(181) = 0.22, *p* = 0.82, [−1.82; 2.29]) and the sex ∗ mean RT interaction (B = −0.003, SEB = 0.02, t (181) = −0.21, *p* = 0.83, [−0.04; 0.03]) did not predict the imitative compatibility effect (see Fig. 14).

When the model included empathy, the model predicted 7.04% of the variance. Empathic concern predicted the imitative compatibility effect above and beyond the base model (B = −0.51, SEB = 0.24, t (179) = −2.08, *p* = 0.04, [−0.99; −0.03]), and explained an addition 2.3% of the variance (ΔR2 = 0.023, F(1,179) = 4.35, p = 0.04; Fig. 15). A decrease in empathic concern predicted a higher imitative compatibility effect. When agreeableness and extraversion were included in the model, the model predicted 7.09% of the variance. Agreeableness marginally predicted the imitative compatibility effect (B = −2.68, SEB = 1.53, t(179) = −1.75, *p* = 0.081, [−5.67; 0.33]) and explained an additional 1.6% of the variance (ΔR2 = 0.0.016, F (1,179) = 3. 07, p = 0.08). Extraversion also marginally predicted the imitative compatibility effect (B = 1.81, SEB = 1.09, t(179) = 1.67, *p* = 0.096, [−0.32; 3.96]) and explained an additional 1.5% of the

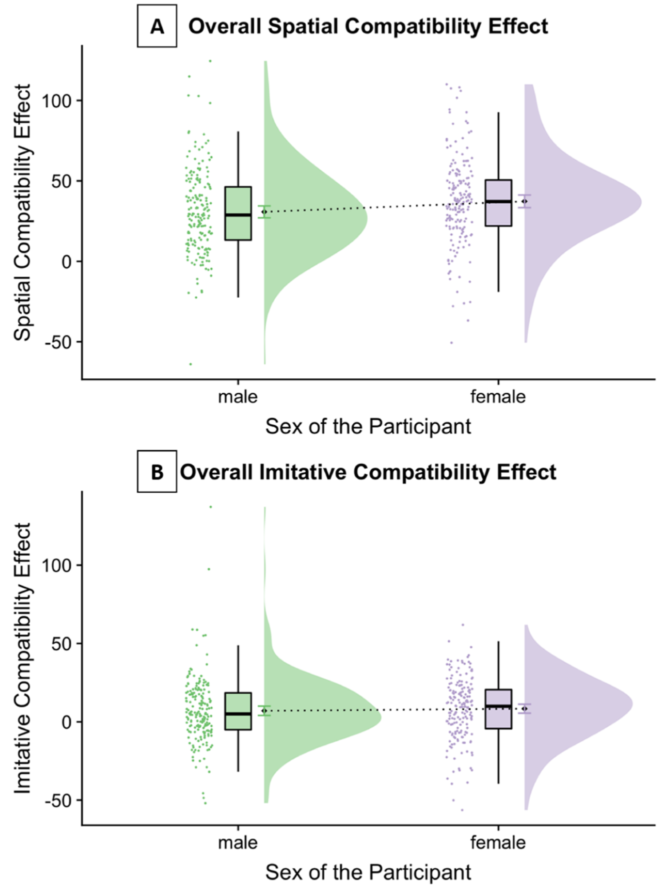


Fig. 13. Experiment 3 – overall compatibility effects. The upper panel (A) shows the spatial compatibility effect collapsed across sex of the stimulus for both males and females. The lower panel (B) shows the imitative compatibility effect collapsed across sex of the stimulus for both males and females. The compatibility effect is calculated by subtracting reaction times on compatible trials from incompatible trials and is measured in milliseconds. Error bars represent 95% confidence intervals.

variance ($\Delta R2 = 0.015$, $F(1,179) = 2.78$, $p = 0.096$). Higher extraversion predicted higher imitative compatibility, whereas higher agreeableness predicted a lower imitative compatibility effect (see Fig. 15). Effect sizes attributable to the addition of empathy (Cohen's $f^2 = 0.03$), and agreeableness and extraversion (Cohen's $f^2 = 0.03$) (beyond the base model) indicated very small effects.

Grandiose and vulnerability narcissism, autistic-like and schizotypal trials, and alexithymia did not predict the imitative compatibility effect above and beyond the base model (all p's > 0.3, all CIs overlapping with zero; see Fig. 14). The multiple regression models are summarized in Supplementary Table 13. Zero-order correlations are also consistent with the findings from the multiple regression analyses (see Supplementary Table 14, Supplementary Fig. 6).

To evaluate the sex ∗ trait interaction terms, we computed additional models – each model consisted of the base model, one trait predictor (subscales were included in the same model), and the sex ∗ trait interaction term. None of the sex ∗ trait interaction terms predicted the compatibility effect above and beyond the base model (Supplementary Fig. 13). Multiple regression models are summarized in Supplementary Table 15. Effect sizes attributable to the addition of the sex ∗ trait interaction terms (beyond the base model) indicated very small effects (Cohen's $f^2 \leq 0.04$ for all models). The pattern of results seen in Experiment 1 for the empathy ∗ sex and alexithymia ∗ sex models did not replicate in Experiment 3 (Fig. 6C).

Although our main question of interest was the link between personality traits and automatic imitation, for completeness, we also report results from the multiple regression analyses for spatial compatibility in the supplementary material (see Supplementary Tables 16 and 17, Supplementary Figs. 14 and 15).

4.3.3.1. Stimuli rating. All participants also rated the male and female hand stimuli on a scale of 1 to 9, with 1 being most masculine, and 9 being most feminine. All male hand stimuli were rated as masculine (Mean rating = 2.93, SD = 0.30). All female stimuli were rated as relatively feminine (Mean rating = 5.68, SD = 0.65). Although the female stimuli were not rated as strongly feminine, the ratings suggest that both male and female stimuli were perceived differently on average by the participants. The stimuli rating data is also available online.

4.4. Discussion

Results from Experiment 3 clearly show that a sex difference exists on the spatial compatibility effect such that females show a higher spatial compatibility effect than males. This difference did not persist when imitative compatibility was measured independently. This suggests that females and males do not differ in the control of automatic imitation as measured by the imitative compatibility effect.

Furthermore, for the first time to date, we manipulated the sex of the stimuli across both male and female participants. Results indicated that there was no own-sex bias in the imitative compatibility effect. For the spatial compatibility effect, although the findings showed a trend toward an own-sex bias such that females showed a greater compatibility effect on female stimuli than male stimuli, this was a relatively small effect size, and thus does not explain much of the sex difference observed in the spatial compatibility effect.

The findings from Experiment 3 thus suggest that it is unlikely that there is a sex difference in the imitative compatibility effect. Instead, our findings suggest that there is a sex difference in the spatial compatibility effect, which may reflect a difference in spatial control between males and females that in the case of this experiment is triggered by the location of a finger in space.

The multiple regression analyses suggest that the imitative compatibility effect is invariant to stable traits of personality including grandiose and vulnerability narcissism, autistic-like and schizotypal traits, as well as alexithymia.

Given prior evidence, we predicted that individuals who report higher empathy, extraversion, and agreeableness would be more prosocial, and would therefore imitate more than those who scored lower on these measures. In the current experiment, although higher extraversion predicted higher imitation, we found the opposite pattern for empathy and agreeableness. An increase in empathic concern and agreeableness predicted a decrease in the imitative compatibility effect. The effects, however, were small and predicted only an additional 2.45% (empathy) and 2.5% (extraversion and agreeableness) of the variance. Before making any firm conclusions, these results would need to be replicated using large sample sizes to ensure that these findings do not reflect false positives. In addition, none of the sex ∗ trait interactions predicted the imitative compatibility effect, and the pattern of results from Experiment 1 for the sex ∗ empathy and sex ∗ alexithymia interactions did not replicate over Experiment 2 and 3 (see Fig. 6). Overall, therefore, these results provide only limited support for small or negligible effects of personality on automatic imitation reinforcing the suggestion that automatic imitation is largely invariant to stable traits of personality (Butler et al., 2015; Cracco et al., 2018).

For all three experiments, we performed all the analyses again by further excluding participants who were three standard deviations away from the group mean on the compatibility effect on either of the tasks. For Experiment 1, no additional participants were excluded. For both Experiment 2 and 3, one additional participant was excluded. Obtained results were very similar to those reported above.

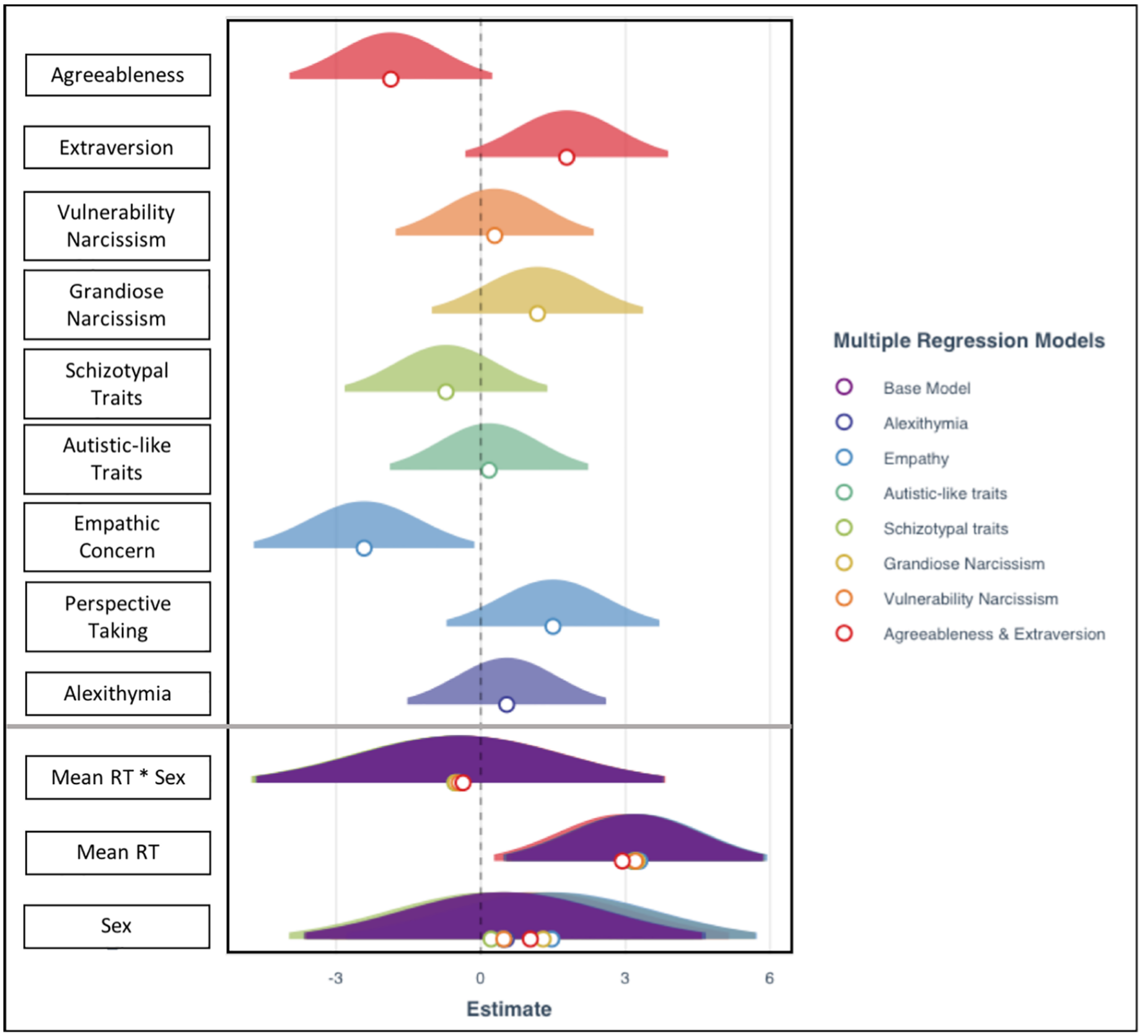


Fig. 14. Experiment 3 – multiple regression analyses. For the outcome variable of imitative compatibility, values of standardised coefficients are plotted for each predictor variable (personality trait) along with their corresponding uncertainties (95% confidence interval width for a normal distribution for each estimate). Coefficients are standardised by dividing by two standard deviation units according to Gelman (2008). The base model consists in the bottom three predictor variables (depicted in violet) – mean RT, Sex, and meanRT ∗ Sex. Abbreviations: RT = Reaction Time. *N.B.* The circles on the purple distributions represent the standardised co-efficients for the main effect of meanRT, main effect of sex, and meanRT ∗ sex interaction respectively for each of the models tested. That is why there are multiple circles for components of the base model because the base model was part of all the models tested (that is, one model for each personality trait). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. General discussion

By integrating methodological approaches from experimental and differential psychology, the current study shines new light on the relationships between stable features of individuals, such as personality and sex, and the architecture of cognitive control systems. Across three experiments, we consistently showed that cognitive control systems are largely invariant to stable aspects of personality, but exhibit a sex difference, such that females show greater interference than males. Moreover, we further qualified this sex difference in two ways. First, we showed that the sex difference was unrelated to the sex of the interaction partner and therefore did not reflect an in-group bias based on sex. Second, we showed that the sex difference was tied to a form of spatial interference control rather than imitative control and therefore it is unlikely to reflect a specialised mechanism for guiding social interactions exclusively. Instead, our findings suggest that a robust sex difference exists in the system (or set of subsystems) that operate in resolving a form of spatial interference control. The implications of these findings for understanding cognitive control systems in social and non-social contexts are discussed.

5.1. Are individual differences in interference control robust and replicable?

In recent years, a key question in psychology and neuroscience has concerned the credibility of reported findings (Button et al., 2013; Munafò et al., 2017; Open Science, 2015; Pashler, Coburn, & Harris, 2012; Vazire, 2018) with estimates of replicability ranging between 25 and 75% (Camerer et al., 2018; Marsman et al., 2017; Matzke et al., 2015; Nosek & Lakens, 2014). Studies that integrate experimental and differential approaches are rare in general, and in the context of imitation control, prior studies have typically used small sample sizes (Ainley et al., 2014; Chartrand & Bargh, 1999; Hogeveen & Obhi, 2013; Obhi et al., 2013; Santiesteban, Bird, Tew, Cioffi, & Banissy, 2015). As such, one important contribution from the current study is a more robust and precise estimate of the size and replicability of sex differences in cognitive control. To do so, we used relatively large sample sizes, which could detect small-to-medium effect sizes with a high degree of confidence, and ran three separate experiments using designs that combined approaches from experimental and differential psychology.

In Experiments 1 and 2, we replicated the sex difference found previously both when the SRC task measured automatic imitation as a

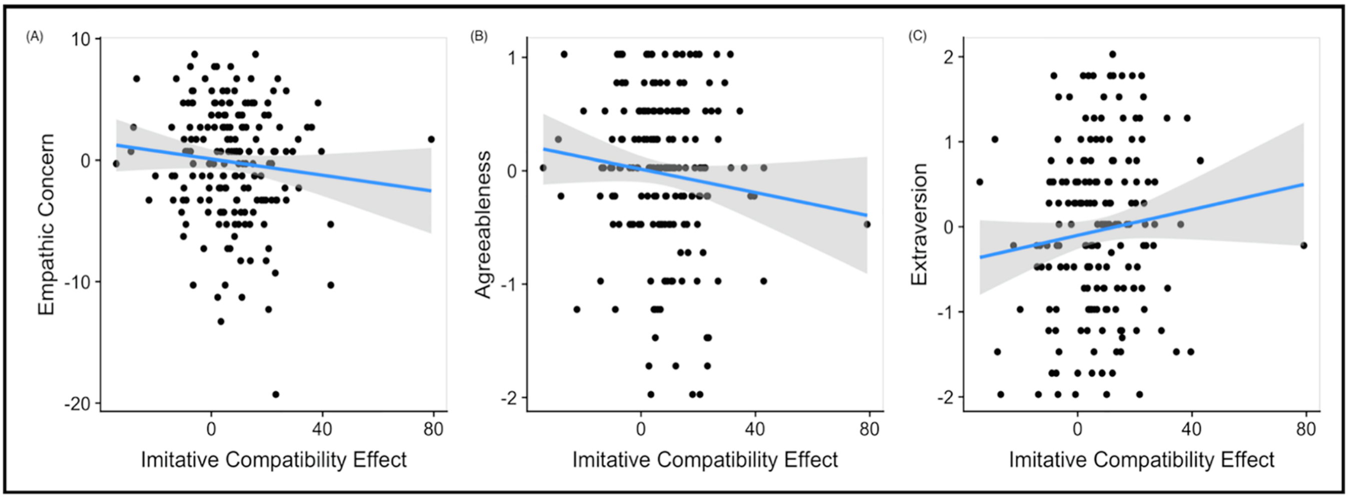


Fig. 15. Experiment 3 – scatterplots. Scatterplots depicting the relationship between imitative compatibility effect and personality traits – empathic concern (A), agreeableness (B), and extraversion (C). X axis denotes the imitative compatibility effect in milliseconds, and Y axis denotes mean centred scores on the personality traits.

composite of imitative and left-right spatial effects (Butler et al., 2015), as well as orthogonal spatial compatibility effects (Genschow et al., 2017). Furthermore, in Experiment 3, we measured imitative compatibility effects independent of spatial compatibility effects (Bertenthal et al., 2006; Boyer et al., 2012; Catmur & Heyes, 2011; Jiménez et al., 2012). In Experiment 3, females showed a greater spatial compatibility effect than males, but there was no difference between the sexes on imitative compatibility. Thus, it is clear that the sex difference on the SRC task reflects a difference in spatial control between males and females, rather than a difference in a specialised system that is dedicated to social control.

According to Cohen's benchmarks for interpreting effect sizes (Cohen, 1992), the difference between the sexes was a small-to-medium effect size (Cohen's d = 0.28) and was relatively consistent across the three experiments, with the lower bound of the 95% CI > 0.02 (see Fig. 16). Considering the sensitivity of our design, it is important to note that these effect sizes were below the 80% power mark, which our power analysis identified, as we had 80% power to detect effects greater than Cohen's d 0.36. Each individual experiment, therefore, had < 80% power to detect a key effect of interest, which surrounded sex differences. This said, all three experiments showed results similar to Butler et al. (2015), in that they were in the same (predicted) direction and of a consistent magnitude even though the effect sizes were small. Further, consistently replicated small effect sizes in large-N studies have the potential to represent the true state of nature more than one-off large effect sizes in single experiment small-N studies. In addition, it is also possible that small effects could gather in their importance over time if the effects aggregate over repeat instances (Funder & Ozer, 2019). More generally, by replicating the effects in separate large sample designs, it makes it less likely that these results represent sampling error (Zwaan et al., 2018). If we interpret the length of confidence interval (Amrhein et al., 2019; Cumming, 2012), then our best estimate is a small to medium effect, with all likely effects being in the predicted direction (i.e., greater than zero). Therefore, building on prior work (Butler et al., 2015), across three large-sample experiments, we have provided a robust and relatively precise estimate of the size of the sex difference and shown that it reflects spatial rather than social control mechanisms.

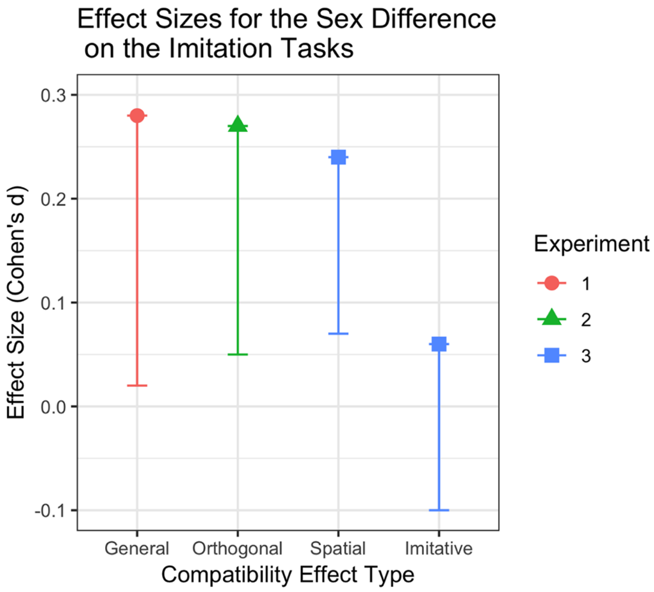


Fig. 16. Effect sizes of the sex difference. Cohen's d effect sizes of the sex difference on the imitation task (compatibility effect) across Experiments 1, 2, and 3. Error bar denotes one-tailed 95% confidence interval.

Moreover, across all three experiments, we consistently found that the control of automatic imitative tendencies, as measured by the SRC task, is invariant to differences in personality traits across individuals. Recently, it has been suggested that the investigation of an experimental effect at the group level, and individual differences within that effect are questions that can be at odds with each other. This is because group effects need low variability within the sample whereas differential psychology questions are dependent on high variability within the sample (Rogosa, 1988). Therefore, we cannot assume that robust experimental paradigms such as the SRC tasks used in the current experiments will lend themselves well to individual difference approaches. Thus, although a difference between groups can be detected if the groups means are sufficiently far away from each other to be detectable, the tasks used may not be able to distinguish between individuals in the population consistently (Hedge, Powell, & Sumner 2018).

Compared to prior studies (Obhi et al., 2013; Hogeveen & Obhi, 2013; Chartrand & Bargh, 1999), however, we provide a more robust test of hypotheses regarding individual differences as we used larger sample sizes, which produce higher statistical power, and we looked for consistent patterns of data across multiple experiments. By doing so, a more stable picture is emerging with regard to personality and SRC measures of automatic imitation, which suggests that mechanisms of imitative control are largely invariant to dimensions of personality (Butler et al., 2015; Cracco et al., 2018; Genschow et al., 2017), even when they are operating in more socially rich contexts (Exp. 1) and when spatial and imitative effects are more clearly separated (Exps. 2 and 3). In short, any effects of personality were small and inconsistent across experiments. Of course, our design did not have sufficient power to detect small effects with reasonable confidence (> 80%), and such effects would require considerably larger sample sizes to be able to confidently confirm that they exist. Thus, our best estimate at present is that the effects of personality on SRC measures of automatic imitation are negligible or small.

5.2. Do differences in cognitive control reflect a sex difference or an in-group bias?

In Experiments 1 and 2, as well as in prior studies that have observed sex differences in the SRC imitation task (Butler et al., 2015; Genschow et al., 2017), the stimuli used were of a female hand. Thus, it was possible that the sex difference reflected an in-group bias leading to higher compatibility effects for females compared to males (Cracco et al., 2018). Indeed, there is already suggestive evidence (from studies with relatively small sample sizes), that both facial imitation and SRC measures of imitation have been found to increase when the interacting partner is an in-group member compared to an out-group member based on race, ethnicity, and arbitrary group assignment (Gleibs et al., 2016; Mondillon et al., 2007; Rauchbauer et al., 2015).

In the current study, based on the sex of the interaction partner, we show no clear evidence for an own-sex bias for either spatial or imitative compatibility. Moreover, in terms of sensitivity, the use of a larger sample size than is typical and 80% power to detect effect sizes at or above Cohen's d = 0.36 means that we can be reasonably confident that effect sizes of this magnitude or larger are unlikely. Taken together, although ingroup biases are potent in everyday life and relate to sex, race and ethnicity (Brown, 1995; Yee & Brown, 1994; Fishbein, 1996, Powlishta, 1995; Rudman & Goodwin, 2004; Kubota et al., 2012; Malpass & Kravitz, 1969), the difference in interference control reported here reflects the sex of the participant, rather than an in-group bias based on the sex of the interaction partner. As such, these results are contrary to proposals put forward by Cracco et al. (2018), and highlight a stable individual difference in interference control, rather than an effect of the social context (i.e., the sex of the interaction partner).

5.3. What type of cognitive system underpins sex differences in interference control?

Three broad structures of cognitive system were candidates to underpin the sex difference in interference control: 1) a sex difference specific to social imitative control; 2) a sex difference generalised across all types of control; 3) a sex difference specific to a form of non-social control. If the sex difference was solely tied to imitative control and reflected the workings of a specialised and domain-specific cognitive structure, we would have observed a sex difference only on the imitative compatibility component of the task. Likewise, if the sex difference reflected the operation of a straightforwardly domain-general system, we would have expected a difference between males and females on the flanker task, as well as both the spatial and imitative components of the automatic imitation SRC task. As such, these findings demonstrate that the sex difference is neither completely domain-general i.e., it does not generalise across all types of compatibility effects nor is it domain-specific i.e., it is not solely tied to the control of automatic imitation.

Our findings show more support for the third type of cognitive system outlined above, which suggests that the sex difference reflects a particular type of non-social interference, which is not shared across all SRC tasks. Indeed, across our experiments, the sex difference was tied to a type of spatial interference observed in the spatial component of the automatic imitation SRC task, but not the imitative component of the same task or the non-social flanker task. A sex difference on spatial control, but not on imitative control, when measured on the same task, suggests that although general cognitive control systems are engaged for both tasks to some extent, they may not be engaged in an identical manner across both the compatibility effects. Moreover, it is unlikely that the sex difference on spatial compatibility reflects a difference in the perceptual processing of the social stimulus (i.e. the hand on the screen) as the stimuli are the same across both compatibility effects, but no sex difference emerges on the imitative compatibility effect. For both imitative and spatial compatibility, therefore, the input to the control mechanism that resolves conflict is the same i.e. a finger. However, the way conflict is resolved for spatial and imitative effects might involve mechanisms that operate differently as a function of sex.

In addition to the sex difference not being tied to social or imitative control, it also reflects a component that is not shared with the flanker task. A lack of sex difference on the flanker task, and little or no correlation between the compatibility effects on the two tasks, has at least two possible interpretations, which are not mutually exclusive. First, it could reflect a lack of sensitivity. The differences between females and males on behavioural indices (such as RT) on the flanker task may be small (Clayson et al., 2011; Fischer, Danielmeier, Villringer, Klein, & Ullsperger, 2016; Stoet, 2010). In the current experiment, our sensitivity analysis suggests that we could detect effect sizes of Cohen's d > 0.36 with reasonable confidence (80%), but the effects of sex on the flanker were smaller than this in Experiment 1 and 2 (Cohen's d = 0.15 and 0.03, respectively). Moreover, a large sample study with 895 participants found a small sex difference in the predicted direction on the flanker task using arrows such that females showed a greater compatibility effects than males (Fischer et al., 2016). Thus, there could also be a non-zero sex difference on the flanker task, but even if this turns out to be the case, it is clear that the size of the sex difference varies across different types of non-social cognitive control tasks.

A second possible reason for the lack of sex difference on the flanker task is that the sex difference is underpinned by a particular type of non-social control. Previous studies help contextualise this finding by showing that females differ from males across a wide range of cognitive control tasks, especially those involving spatial processing (Bayliss et al., 2005; Clayson et al., 2011; Stoet, 2017; Stoet, 2010). One possibility, therefore, is that the sex difference may reflect a difference in the two types of spatial conflict measured by the flanker and spatial compatibility effect. For example, in the SRC task measuring spatial compatibility, the conflict arises because a stimulus feature is inconsistent to the response, whereas the flanker task measures both stimulus-stimulus and stimulus-response conflicts (Kornblum, 1994; Kornblum, Hasbroucq, & Osman, 1990; Verbruggen, Notebaert, Liefooghe, & Vandierendonck, 2006). It has been proposed that stimulus-response (S-R) conflicts and stimulus-stimulus (S-S) conflicts are underpinned by different processing patterns (Frühholz, Godde, Finke, & Herrmann, 2011; Kornblum et al., 1990; Kornblum & Lee, 1995; Kornblum, Stevens, Whipple, & Requin, 1999; Li, Nan, Wang, & Liu, 2014; Zhang, Zhang, & Kornblum, 1999). Therefore, these types of conflicts would be worth taking into consideration in future research that investigates individual differences in automatic imitation, and social and non-social cognitive control.

More generally, other sex differences, which do not rely on SRC paradigms, can further contextualise our findings. Indeed, prior research suggests that females differ from males on a range of social processes (Baron-Cohen, 2002). For example, females show greater empathy than males, which may lead to more pro-social behaviour, thus suggesting that females may imitate more than males (Baron-Cohen & Wheelwright, 2004; Christov-Moore et al., 2014; Schulte-Rüther, Markowitsch, Shah, Fink, & Piefke, 2008). However, although empathy has been associated with a variety of paradigms investigating automatic imitation (Chartrand & Bargh, 1999; Müller et al., 2013; Sonnby-Borgström, 2002), there does not seem to be a clear link between empathy and automatic imitation as measured on the SRC task (Butler et al., 2015; Genschow et al., 2017). Moreover, while females show higher facial mimicry than males (Sonnby-Borgström, 2002; Sonnby-Borgström et al., 2008; Dimberg, 1990; Hess & Bourgeois, 2010; Korb et al., 2015; Lundqvist, 1995), studies investigating imitation of other behaviours, such as nose-scratching, have not found any reliable sex differences, although such studies have been limited by small sample sizes (Chartrand & Bargh, 1999). Inconsistent and equivocal results across imitation tasks might suggest that these tasks engage different cognitive mechanisms.

In addition, while we separated the spatial and imitative effects in Experiment 3 using a modified version of the paradigm developed by Catmur and Heyes (2011), there are other tasks that measure imitative effects that are not confounded with spatial compatibility effects (e.g. Bortoletto, Mattingley, & Cunnington, 2013). Thus, any conclusions we make are limited to the type of task we used in the current experiment. However, whether other tasks are more powerful measures of imitative effects is an empirical question that would need to be tested by future research.

These findings raise a fundamental issue that remains unresolved: does the SRC imitation task actually link to imitation that occurs in more naturalistic real-life settings, and social cognition more broadly? The domain of social cognition research has seen the use of cognitive psychology paradigms and methodologies to answer questions that are of interest to social psychologists (Lambert & Scherer, 2013). But are these paradigms truly measuring what we think they are measuring? More and more recent evidence suggests that social cognition is fundamentally different when we are involved in live social interactions with each other as compared to when we are doing tasks in a controlled environment (Redcay & Schilbach, 2019; Schilbach et al., 2013). For instance, researchers have suggested that eye-gaze behaviours when measured using screen-based tasks cannot be validly generalised to and used as a proxy for understanding gaze behaviours in live social interaction settings (Grossman, Zane, Mertens, & Mitchell, 2019). Thus, more empirical evidence is needed to know whether the SRC task of automatic imitation is actually measuring social cognitive processes that are specifically tied to imitation, rather than more domain-general cognitive control processes such as selection and prioritisation, and whether any such processes relate to imitation "in the wild" (Ramsey, 2018). Finally, it is worth noting that imitation is likely to rely on many different cognitive components and each imitation task may map on to different combinations of these processes. Indeed, although different measures of automatic imitation have been previously assumed to rely on the same underlying mechanisms, there is accumulating empirical and theoretical reason to question such an assumption (Genschow et al., 2017; Ramsey, 2018). Thus, divergent sex differences across measures of different dimensions of automatic imitative behaviour may reflect differences in cognitive mechanisms that underpin these tasks.

5.4. Limitations and constraints on generality

In the current work, we make conclusions about social and non-social cognitive control on the basis of compatibility effects measured on flanker and imitation tasks. This makes sense because we had predictions specifically about individual differences in social and non-social cognitive control as measured by compatibility effects. However, there are many processes that contribute to such compatibility effects, and cognitive control is only one of them. Therefore, we cannot be sure that individual differences (or lack thereof) on these tasks are tied to individual differences in cognitive control (Musslick, Cohen, & Shenhav, 2019). Thus, an important and complementary direction for future research would be to disentangle the extent to which inter-individual differences reflect differences in cognitive control or other cognitive capacities. To do so, one may include a baseline condition for both the flanker and imitation tasks, which may help to separate differences in cognitive control from differences in other factors that also contribute to SRC tasks.

A second potential limitation to the current work is the use of the flanker task as a measure of non-social cognitive control. Cognitive control has various components which can be measured by many different tasks (Rondeel, Van Steenbergen, Holland, & van Knippenberg, 2015). However, we use different non-social tasks across the three experiments (i.e. flanker in Experiments 1 and 2, and spatial compatibility in Experiment 3) as a comparison with the automatic imitation task and find similar results irrespective of the type of task we used. Therefore, although every task has drawbacks, we feel that our key findings are relatively robust in the sense that they remain largely indifferent to changes in experimental design.

5.5. Conclusion

The current findings provide a general insight into the relationship between individual differences and cognitive control systems in social and non-social contexts. Integrating experimental and differential psychology approaches, across three large sample experiments, we show that there is negligible or no evidence for a link between social control and stable personality traits. However, cognitive control systems vary as a function of biological sex, such that females show a greater interference than males. Further, this sex difference does not reflect an in-group bias based on the sex of the interacting partner, and is not tied specifically to social control but reflects differences in the cognitive systems that operate in resolving a form of spatial interference. Therefore, we show that the sex difference exists in the system (or set of subsystems) that operate in resolving a form of spatial interference control, and that such systems are unaffected by social factors such as facial expression or the sex of the interaction partner. More generally, the results highlight the value of integrating approaches from experimental and differential psychology, as well as using large sample sizes, in order to investigate the relationship between cognitive control architectures and stable traits of individuals, which few studies have achieved to date.

CRediT authorship contribution statement

Kohinoor M. Darda: Conceptualization, Methodology, Investigation, Formal analysis, Writing - original draft, Visualization. **Emily E. Butler**: Conceptualization, Methodology, Investigation. **Richard Ramsey**: Conceptualization, Methodology, Writing - review & editing, Supervision.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104317.

References

Aboud, F. E. (1988). *Children and prejudice.* Oxford: Basil Blackwell.

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology, 60*, 693–716.

Ainley, V., Brass, M., & Tsakiris, M. (2014). Heartfelt imitation: High interoceptive awareness is linked to greater automatic imitation. *Neuropsychologia, 60*, 21–28.

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2018). Raincloud plots: A multi-platform tool for robust data visualization. *PeerJ Preprints, 6*, e27137v1.

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief “red flags” for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(2), 202–212.

Allport, G. W. (1954). *The nature of prejudice.* Cambridge/Reading, MA: Addison-Wesley.

Alwall, N., Johansson, D., & Hansen, S. (2010). The gender difference in gaze-cueing: Associations with empathizing and systemizing. *Personality and Individual Differences, 49*(7), 729–732.

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality, 40*. https://doi.org/10.1016/j.jrp.2005.03.002 (440-350).

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A.

E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science, 15*(2), 88–93.

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*, 305–307.

Ashton, M. C., Paunonen, S. V., Helmes, E., & Jackson, D. N. (1998). Kin altruism, reciprocal altruism, and the Big Five personality factors. *Evolution and Human Behavior, 19*(4), 243–255.

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23–32.

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science, 18*, 89–94.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences, 6*, 248–254.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*, 163–175.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*, 5–17.

Barrio, V. D., Aluja, A., & García, L. F. (2004). Relationship between empathy and the Big Five personality traits in a sample of Spanish adolescents. *Social Behavior and Personality: An International Journal, 32*, 677–681. https://doi.org/10.2224/sbp.2004.32.7.677.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest, 20*(1), 1–68.

Bayliss, A. P., di Pellegrino, G., & Tipper, S. P. (2005). Sex differences in eye gaze and symbolic cueing of attention. *Quarterly Journal of Experimental Psychology, 58A*, 631–650.

Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior, 12*, 120–138.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science, 18*(8), 706–712.

Bertenthal, B. I., Longo, M. R., & Kosobud, A. (2006). Imitative response tendencies following observation of intransitive actions. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 210.

Binney, J. R., & Ramsey, R. (2020). Social Semantics: The role of conceptual knowledge and cognitive control in a neurobiological model of the social brain. *Neuroscience and Biobehavioural Reviews, 12*, 28–38.

Bortoletto, M., Mattingley, J. B., & Cunnington, R. (2013). Effects of Context on Visuomotor Interference Depends on the Perspective of Observed Actions. *Plos One, 8*(1), e53248.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624.

Botvinick, M. M., & Cohen, J. D. (2014). The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science, 38*(6), 1249–1285.

Boyer, T. W., Longo, M. R., & Bertenthal, B. I. (2012). Is automatic imitation a specialized form of stimulus–response compatibility? Dissociating imitative and spatial compatibilities. *Acta Psychologica, 139*, 440–448.

Brass, M., Bekkering, H., Wohlschlager, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial, and imitative cues. *Brain and Cognition, 44*(2), 124–143. 11041986.

Brass, M., & Heyes, C. (2005). Imitation: Is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences, 9*(10), 489–495. https://doi.org/10.1016/j.tics.2005.08.007.

Brown, R. (1995). *Prejudice. Its social psychology.* Oxford, U.K.: Blackwell.

Butler, E. E., Ward, R., & Ramsey, R. (2015). Investigating the relationship between stable personality characteristics and automatic imitation. *PLoS One, 10*(6), e0129651. https://doi.org/10.1371/journal. pone.012965.

Butler, E. E., Ward, R., & Ramsey, R. (2016). The influence of facial signals on the automatic imitation of hand actions. *Frontiers in Psychology, 7*(1653), https://doi.org/10.3389/fpsyg.2016.01653.

Button, K. S., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience, 14*, 365–376.

Cacioppo, J. T., Berntson, G. G., Sheridan, J. F., & McClintock, M. K. (2000). Multi-level integrative analyses of human behavior: Social neuroscience and the complementing nature of social and biological approaches. *Psychological Bulletin, 126*, 829–843.

Camerer, C. F., Dreber, A., Holzmeister, H., H, T., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behavior, 2*, 637–644.

Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2001). Children's lay theories about ingroups and outgroups: Reconceptualizing research on prejudice. *Personality and Social Psychology Review, 5*(2), 118–128.

Campbell, R., Elgar, K., Kuntsi, J., Akers, R., Terstegge, J., Coleman, M., & Skuse, D. (2002). The classification of "fear" from faces is associated with face recognition skill in females. *Neuropsychologia, 40*(6), 575–584. https://doi.org/10.1016/S0028-3932(01)00164-6.

Carling, K. (2000). Resistant outlier rules and the non-Gaussian case. *Computational Statistics & Data Analysis, 33*, 249–258.

Catmur, C., & Heyes, C. (2011). Time course analyses confirm independence of imitative and spatial compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 37*(2), 409–421.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. Guilford Press.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893–910.

Cho, Y. S., & Proctor, R. W. (2003). Stimulus and response representations underlying orthogonal stimulus-response compatibility effects. *Psychonomic Bulletin & Review, 10*(1), 45–73.

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews, 46*, 604–627.

Clayson, P. E., Clawson, A., & Larson, M. J. (2011). Sex differences in electrophysiological indices of conflict monitoring. *Biological Psychology, 87*, 282–289.

Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology, 112*, 155–159.

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., ... Brass, M. (2018). Automatic imitation: A meta-analysis. *Psychological Bulletin, 144*(5), 453.

Cramer, A. O., Waldorp, L. J., Van Der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioural and brain sciences, 33*(2–3), 137–150.

Crescentini, C., Mengotti, P., Grecucci, A., & Rumiati, R. I. (2011). The effect of observed biological and non biological movements on action imitation: An fMRI study. *Brain Research, 1420*, 80–92. https://doi.org/10.1016/j.brainres.2011.08.077.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*(11), 671.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*(2), 116.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York: Routledge.

Darda, K. M., Butler, E. E., & Ramsey, R. (2018). Functional specificity and sex differences in the neural circuits supporting the inhibition of automatic imitation. *Journal of Cognitive Neuroscience, 30*(6), 914–933.

Darda, K. M., & Ramsey, R. (2019). The inhibition of automatic imitation: A meta-analysis and synthesis of fMRI studies. *NeuroImage, 197*, 320–329.

Davis, M. H. (1980). Interpersonal reactivity index (IRI). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

de Schotten, M. T., & Shallice, T. (2017). Identical, similar or different? Is a single brain model sufficient? *PLoS One, 86*, 172–175. https://doi.org/10.1016/j.cortex.2016.12.002.

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology, 19*, 643–647.

Dimberg, U. (1990). Gender differences in facial reactions to facial expressions. *Biological Psychology, 30*, 151–159.

Dodell-Feder, D., Ressler, K. J., & Germine, L. T. (2019). Social cognition or social class and culture? On the interpretation of differences in social cognitive performance. *Psychological Medicine,* 1–13.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment, 18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149.

Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology, 73*(6), 1224.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach.* New York: Plenum Press.

Fischer, A. G., Danielmeier, C., Villringer, A., Klein, T. A., & Ullsperger, M. (2016). Gender influences on brain responses to errors and post-error adjustments. *Scientific Reports, 6*, 24435.

Fischer-Baum, S., Kook, J. H., Lee, Y., Ramos Nuñez, A. I., & Vannucci, M. (2018). Individual differences in the neural and cognitive mechanisms of single word reading. *Frontiers in Human Neuroscience, 12*, 271.

Fishbein, H. D. (1996). *Peer prejudice and discrimination: Evolutionary, cultural, and developmental dynamics.* Boulder, CO: Westview Press.

Frith, C. D. (2008). Social cognition. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 363*(1499), 2033–2039.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63*, 287–313.

Frühholz, S., Godde, B., Finke, M., & Herrmann, M. (2011). Spatio-temporal brain dynamics in a combined stimulus–stimulus and stimulus–response conflict task. *Neuroimage, 54*(1), 622–634.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168.

Geary, D. C. (2010). *Male, female: The evolution of human sex differences.* American Psychological Association.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics, 2*(4), 1360–1383.

Genschow, O., van Den Bossche, S., Cracco, E., Bardi, L., Rigoni, D., & Brass, M. (2017). Mimicry and automatic imitation are not correlated. *PLoS One, 12*, e0183784.

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*2515245918771329.

Gleibs, I. H., Wilson, N., Reddy, G., & Catmur, C. (2016). Group dynamics in automatic imitation. *PLoS One, 11*(9), e0162880.

Gowen, E., Bolton, E., & Poliakoff, E. (2016). Believe it or not: Moving non-biological stimuli believed to have human origin can be represented as human movement. *Cognition, 146*, 431–438.

Graziano, W. G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality. *Handbook of personality psychology* (pp. 795–824). Academic Press.

Grecucci, A., Koch, I., & Rumiati, R. I. (2011). The role of emotional context in facilitating imitative actions. *Acta Psychologica, 138*(2), 311e315. https://doi.org/10.1016/j.actpsy.2011.07.005.

Grossman, R. B., Zane, E., Mertens, J., & Mitchell, T. (2019). Facetime vs. screentime: Gaze patterns to live and video social stimuli in adolescents with ASD. *Scientific Reports, 9*(1), 1–10.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85*(4), 845.

Hamilton, A. F.d. C. (2013). Reflecting on the mirror neuron system in autism: A systematic review of current theories. *Developmental Cognitive Neuroscience, 3*, 91–105. https://doi.org/10.1016/j.dcn.2012.09.008.

Hansen, J., Alves, H., & Trope, Y. (2016). Psychological distance reduces literal imitation: Evidence from an imitation-learning paradigm. *Journal of Experimental Psychology: Human Perception and Performance,* 320–330. https://doi.org/10.1037/xhp0000150.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behaviour Research Methods, 50*(3), 1166–1186.

Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's Narcism Scale. *Journal of Research in Personality, 31*(4), 588–599.

Hess, U., & Bourgeois, P. (2010). You smile–I smile: Emotion expression in social interaction. *Biological Psychology, 84*(3), 514–520.

Heyes, C. (2009). Evolution, development and intentional control of imitation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 364*(1528), 2293–2298. https://doi.org/10.1098/rstb. 2009.0049.

Heyes, C. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463.

Hogeveen, J., & Obhi, S. S. (2013). Automatic imitation is automatic, but less so for narcissists. *Experimental Brain Research, 224*, 613–621. https://doi.org/10.1007/s00221-012-3339-6.

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65*, 373–398 (Crossref, Google Scholar).

Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends in Cognitive Sciences, 19*(3), 126–132.

Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in Cognitive Sciences, 13*(12), 524–531.

JASP Team (2018). *JASP (version 0.9).* (Computer software).

Jeffreys, H. (1961). *Theory of probability.* Oxford: UK: Oxford University Press.

Jiménez, L., Recio, S., Méndez, A., Lorda, M. J., Permuy, B., & Méndez, C. (2012). Automatic imitation and spatial compatibility in a key-pressing task. *Acta Psychologica, 141*(1), 96–103.

Judge, J., & Taylor, P. J. (2012). Gender differences on the semantic flanker task using transposed-letter target words. *The Quarterly Journal of Experimental Psychology, 65*(10), 2008–2017.

Kao, L. S., & Green, C. E. (2008). Analysis of variance: Is there a difference in means and what does it mean? *Journal of Surgical Research, 144*(1), 158–170.

Kavanagh, L. C., & Winkielman, P. (2016). The functionality of spontaneous mimicry and its influences on affiliation: An implicit socialization account. *Frontiers in Psychology, 7*, 458.

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology, 13*(6), 522–525.

Korb, S., Malsert, J., Rochas, V., Rihs, T. A., Rieger, S. W., Schwab, S., ... Grandjean, D. (2015). Gender differences in the neural network of facial mimicry of smiles–An rTMS study. *Cortex, 70*, 101–114.

Kornblum, S. (1994). The way irrelevant dimensions are processeddepends on what they overlap with: The case of Stroop- and Simon-like stimuli. *Psychological Research/ Psychologische For-schung, 56*, 130–135.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility—A model and taxonomy. *Psychological Review, 97*(2), 253.

Kornblum, S., & Lee, J. W. (1995). Stimulus-response compatibility with relevant and irrelevant stimulus dimensions that do and do not overlap with the response. *Journal of Experimental Psychology: Human Perception and Performance, 21*(4), 855.

Kornblum, S., Stevens, G. T., Whipple, A., & Requin, J. (1999). The effects of irrelevant stimuli: 1. The time course of stimulus–stimulus and stimulus–response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of Experimental Psychology: Human Perception and Performance, 25*(3), 688.

Krach, S., Blumel, I., Marjoram, D., Lataster, T., Krabbendam, L., & Weber, J. (2009). Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. *BMC Neuroscience, 10*, 9.

Kubota, J. T., Banaji, M. R., & Phelps, E. A. (2012). The neuroscience of race. *Nature Neuroscience, 15*(7), 940.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863.

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science, 14*(4), 334e339. https://doi.org/10.1111/1467-9280.14481.

Lambert, A. J., & Scherer, L. (2013). Measurement and methodology in social cognition: A historical perspective. *The Oxford handbook of social cognition. 33*.

Larsen, H., Overbeek, G., Granic, I., & Engels, R. C. M. E. (2010). Imitation of alcohol consumption in same-sex and other-sex dyads. *Alcohol and Alcoholism, 45*(6), 557–562. https://doi.org/10.1093/1lcalc/agq053.

Levant, R. F., Hall, R. J., Williams, C. M., & Hasan, N. T. (2009). Gender differences in alexithymia. *Psychology of Men & Masculinity, 10*(3), 190.

Li, Q., Nan, W., Wang, K., & Liu, X. (2014). Independent processing of stimulus-stimulus and stimulus-response conflicts. *PLoS One, 9*(2), e89249.

Lundqvist, L. O. (1995). Facial EMG reactions to facial expressions: A case of facial emotional contagion? *Scandinavian Journal of Psychology, 36*(2), 130–141.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other "race". *Journal of Personality and Social Psychology, 13*, 330–334.

Marsh, L. E., Bird, G., & Catmur, C. (2016). The imitation game: Effects of social cues on "imitation" are domain-general in nature. *Neuroimage, 139*, 368–375.

Marsman, M., Schönbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., & Wagenmakers, E. J. (2017). A Bayesian bird's eye view of "replications of important results in social psychology". *Royal Society Open Science, 4*(1), 160426.

Matzke, D., Nieuwenhuis, S., van Rijn, H., Slagter, H. A., van der Molen, M. W., & Wagenmakers, E.-J. (2015). The effect of horizontal eye movements on free recall: A preregistered adversarial collaboration. *Journal of Experimental Psychology: General, 144*(1), e1–15.

McCrae, R. R., & Costa, P. T., Jr. (1999). A five-factor theory of personality. *Handbook of personality: Theory and research. 2(1999). Handbook of personality: Theory and research* (pp. 139–153).

Merritt, P., Hirshman, E., Wharton, W., Stangl, B., Devlin, J., & Lenz, A. (2007). Evidence for gender differences in visual selective attention. *Personality and Individual Differences, 43*(3), 597–609.

Michael, J., & D'Ausilio, A. (2015). Domain-specific and domain-general processes in social perception–A complementary approach. *Consciousness and Cognition, 36*, 434–437.

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18*(1), 37–45. https://doi.org/10.1016/j.tics.2013.10.011.

Milner, D. (1983). *Children and race: Ten years on.* London: Ward Lock Educational.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100.

Mondillon, L., Niedenthal, P. M., Gil, S., & Droit-Volet, S. (2007). Imitation of in-group versus outgroup members' facial expressions of anger: A test with a time perception task. *Social Neuroscience, 2*(3–4), 223–237. https://doi.org/10.1080/17470910701376894.

Müller, B. C., Leeuwen, M. L., Baaren, R. B., Bekkering, H., & Dijksterhuis, A. (2013). Empathy is a beautiful thing: Empathy predicts imitation only for attractive others. *Scandinavian Journal of Psychology, 54*, 401–406.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., ... Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 21. https://doi.org/10.1038/s41562-016-0021.

Musslick, S., Cohen, J. D., & Shenhav, A. (2019). *Decomposing individual differences in cognitive control: A model-based approach. Proceedings of the 41st annual meeting of the Cognitive Science Society*Montreal, CA: Cognitive Science Society2427–2433.

Nosek, B. A., & Lakens, D. (2014). *Registered reports*.

Oberman, L. M., & Ramachandran, V. S. (2007). The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin, 133*(2), 310–327. https://doi.org/10.1037/0033-2909.133.2.310.

Obhi, S. S., Hogeveen, J., Giacomin, M., & Jordan, C. H. (2013). Automatic imitation is reduced in narcissists. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 920–928. https://doi.org/10.1037/a0034056.

Ochsner, K. N., & Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist, 56*(9), 717.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, 6251.

O'Reilly, R. C., Herd, S. A., & Pauli, W. M. (2010). Computational models of cognitive control. *Current Opinion in Neurobiology, 20*, 257–261.

Parsons, S., Kruijt, A. W., & Fox, E. (2019). Psychological Science Needs a Standard Practice of Reporting the Reliability of Cognitive-Behavioral Measurements. *Advances in Methods and Practices in Psychological Science, 2*(4), 378–395.

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS One, 7*(8), e42510. https://doi.org/10.1371/journal.pone.0042510.

Payne, B. K. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology, 89*(4), 488.

Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust correlation analyses: False positive and power validation using a new open source Matlab toolbox. *Frontiers in Psychology, 3*, 606. https://doi.org/10.3389/fpsyg.2012.00606.

Powlishta, K. K. (1995). Intergroup processes in childhood: Social categorization and sex role development. *Developmental Psychology, 31*, 781–788.

Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Sex, sexual orientation, and identification of positive and negative facial affect. *Brain and Cognition, 54*, 179–185. https://doi.org/10.1016/j.bandc.2004.01.002.

Raine, A., & Benishay, D. (1995). The SPQ-B: A brief screening instrument for schizotypal personality disorder. *Journal of Personality Disorders, 9*(4), 346–355.

Ramsey, R. (2018). What are reaction time indices of automatic imitation measuring? *Consciousness and Cognition, 65*, 240–254.

Ramsey, R., & Ward, R. (in press). Putting the non-social into social neuroscience: A role for domain-general priority maps during social interactions. Perspectives on Psychological Science. preprint: https://doi.org/10.31234/osf.io/tqwfn.

Rauchbauer, B., Majdandžić, J., Hummer, A., Windischberger, C., & Lamm, C. (2015). Distinct neural processes are engaged in the modulation of mimicry by social group-membership and emotional expressions. *Cortex, 70*, 49–67. https://doi.org/10.1016/j.cortex.2015.03.007.

Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus. *Developmental Science, 14*(1), 92–105.

R Core Team (2018). *R: A language and environment for ## statistical computing.* Vienna, Austria: R Foundation for Statistical Computing https://www.R-project.org/.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience, 20*(8), 495–505.

Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.). *Methodological issues in ageing research* (pp. 171–210). New York: Springer.

Rondeel, E., Van Steenbergen, H., Holland, R., & van Knippenberg, A. (2015). A closer look at cognitive control: Differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in Human Neuroscience, 9*, 494.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association, 79*, 871–881.

Rousseeuw, P. J., & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics, 41*, 212–223.

Rubia, K., Hyde, Z., Halari, R., Giampietro, V., & Smith, A. (2010). Effects of age and sex on developmental neural networks of visual-spatial attention allocation. *Neuroimage, 51*, 817–827.

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology, 87*(4), 494.

Russell, T. A., Tchanturia, K., Rahman, Q., & Schmidt, U. (2007). Sex differences in theory of mind: A male advantage on Happé's "cartoon" task. *Cognition and Emotion, 21*, 1554–1564. https://doi.org/10.1080/02699930601117096.

Santiesteban, I., Bird, G., Tew, O., Cioffi, M. C., & Banissy, M. J. (2015). Mirror-touch synaesthesia: Difficulties inhibiting the other. *Cortex, 71*, 116–121. https://doi.org/10.1016/j.cortex.2015.06.019.

Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., & Bird, G. (2012). Training social cognition: From imitation to theory of mind. *Cognition, 122*(2), 228–235.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and Brain Sciences, 36*(4), 393–414.

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in big five personality traits across 55 cultures. *Journal of Personality and Social Psychology, 94*(1), 168–182. https://doi.org/10.1037/0022-3514.94.1.168.

Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *Neuroimage, 42*(1), 393–403.

Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental Science, 13*(4), 599–610.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. A. (2012). 21 Word Solution. Available at SSRN: https://ssrn.com/abstract=2160588 or doi: 10.2139/ssrn.2160588.

Sonnby-Borgström, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scandinavian Journal of Psychology, 43*, 433–443.

Sonnby-Borgström, M., Jönsson, P., & Svensson, O. (2008). Gender differences in facial imitation and verbally reported emotional contagion from spontaneous to emotionally regulated processing levels. *Scandinavian Journal of Psychology, 49*, 111–122. https://doi.org/10.1111/j.1467-9450.2008.00626.x.

Southgate, V., & Hamilton, A. F.d. C. (2008). Unbroken mirrors: Challenging a theory of autism. *Trends in Cognitive Sciences, 12*(6), 225–229. https://doi.org/10.1016/j.tics.2008.03.005.

Sowden, S., Brewer, R., Catmur, C., & Bird, G. (2016). The specificity of the link between alexithymia, interoception, and imitation. *Journal of Experimental Psychology: Human Perception and Performance, 42*(11), 1687.

Spunt, R. P., & Adolphs, R. (2017). A new look at domain specificity: Insights from social neuroscience. *Nature Reviews Neuroscience, 18*, 559.

Stoet, G. (2010). Sex differences in the processing of flankers. *The Quarterly Journal of Experimental Psychology, 63*, 633–638.

Stoet, G. (2017). Sex differences in the Simon task help to interpret sex differences in selective attention. *Psychological Research, 81*, 571–581.

Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1746–1759. https://doi.org/10.1037/0096-1523.26.6.1746.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178.

Thakkar, K. N., Peterman, J. S., & Park, S. (2014). Altered brain activation during action imitation and observation in schizophrenia: A translational approach to investigating social dysfunction in schizophrenia. *American Journal of Psychiatry, 171*(5), 539–548.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., ... Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249.

Trujillo-Ortiz, A., Hernandez-Walls, R., Barba-Rojo, K., & Cupul-Magana, L. (2007). *HZmvntest:Henze-Zirkler's multivariate normality test. A MATLAB file*.

van Baaren, R., Janssen, L., Chartrand, T. L., & Dijksterhuis, A. (2009). Where is the love? The social aspects of mimicry. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 364*(1528), 2381–2389. https://doi.org/10.1098/rstb.2009.0057.

Van Baaren, R. B., Holland, R. W., Steenaert, B., & van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology, 39*(4), 393–398.

Van Bavel, J. J., & Cunningham, W. A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin, 35*(3), 321–335.

Van der Graaff, J., Branje, S., De Wied, M., Hawk, S., Van Lier, P., & Meeus, W. (2014). Perspective taking and empathic concern in adolescence: Gender differences in developmental changes. *Developmental Psychology, 50*(3), 881.

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science, 13*(4), 411–417.

Verboten, S., & Hubert, M. (2005). LIBRA: A MATLAB library for robust analysis. *Chemometrics and Intelligent Laboratory Systems,* (75), 127–136.

Verbruggen, F., Notebaert, W., Liefooghe, B., & Vandierendonck, A. (2006). Stimulus-and response-conflict-induced cognitive control in the flanker task. *Psychonomic Bulletin & Review, 13*(2), 328–333.

Wang, Y., & Hamilton, A. F.d. C. (2014). Why does gaze enhance mimicry? Placing gaze-mimicry effects in relation to other gaze phenomena. *Quarterly Journal of Experimental Psychology, 67*(4), 747–762. https://doi.org/10.1080/17470218.2013.828316.

Wang, Y., Ramsey, R., & Hamilton, A. F. D. C. (2011). The control of mimicry by eye contact is mediated by medial prefrontal cortex. *Journal of Neuroscience, 31*(33), 12001–12010.

Webb, J. T. (1972). Interview synchrony: An investigation of two speech rate measures in an automated standardized interview. In B. Pope, & A. W. Siegman (Eds.). *Studies in dyadic communication* (pp. 115–133). New York: Pergamon.

Williams, J. H., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience and Biobehaviour Review, 25*, 287–295. https://doi.org/10.1016/S0149-7634(01)00014-8.

Weeks, D. J., & Proctor, R. W. (1990). Salient-features coding in the translation between orthogonal stimulus and response dimensions. *Journal of Experimental Psychology: General, 119*(4), 355–366. https://doi.org/10.1037//0096-3445.119.4.355.

Weeks, D. J., Proctor, R. W., & Beyak, B. (1995). Stimulus–response compatibility for vertically oriented stimuli and horizontally oriented responses: Evidence for spatial coding. *The Quarterly Journal of Experimental Psychology, 48*(2), 367–383.

Yee, M., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology, 33*, 183–196.

Zhang, H. H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of stimulus–stimulus and stimulus–response compatibility. *Cognitive Psychology, 38*(3), 386–432.

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences, 41*.