



MACQUARIE
University

Macquarie University PURE Research Management System

This is the Accepted Manuscript version of the following article:

Darda, K. M., Butler, E. E., & Ramsey, R. (2020). Individual differences in social and non-social cognitive control. *Cognition*, 202, 104317.

which has been published in final form at:

<https://doi.org/10.1016/j.cognition.2020.104317>

© 2020. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

1 **Journal:**

2

3

4

5

6

7

8 **Title: Individual Differences in Social and Non-Social Cognitive Control**

9 **Kohinoor M. Darda^{1,2}, Emily E. Butler¹ and Richard Ramsey²**

10

11 ¹ Wales Institute for Cognitive Neuroscience, School of Psychology, Bangor University,
12 Bangor, Gwynedd, Wales, LL57 2AS, United Kingdom

13 ² Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia.

14

15 Corresponding author: richard.ramsey@mq.edu.au

16

17 Word count: 20657

18 Key words: cognitive control; sex differences; personality; automatic imitation.

19

20

Abstract

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

45

General Introduction

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

63 Cognitive control is multi-faceted, with a core function being the ability to inhibit
64 unwanted but dominant responses, in order to prioritise alternative, context-
65 appropriate responses (Miyake et al., 2000; Payne, 2005; Chaiken & Trope, 1999). For
66 example, non-social cognitive control may involve inhibiting automatic reading
67 responses in a Stroop task (MacLeod, 1991), whereas social cognitive control may
68 involve controlling automatic social biases based on race, sex or other social groupings

69 (Amodio et al., 2004). The study of cognitive processes during social interactions has
70 received much attention in the last couple of decades across different methodologies
71 (Adolphs, 2009; Ochsner & Lieberman, 2001; Frith, 2008; Frith & Frith, 2012). For
72 example, researchers investigating social cognition have used methodologies from
73 cognitive psychology to provide experimental control over phenomena of interest to
74 social psychologists (Lambert & Scherer, 2013). One such example is that of
75 methodologies used to study automatic imitation.

76 Humans imitate a wide range of behaviours from their interaction partners,
77 including speech patterns, body postures, gestures and facial expressions (Brass et al.,
78 2000; Hansen et al., 2016; Bernieri, 1988; Dimberg, 1982; Webb, 1972). This behaviour
79 is usually not intended, often occurs without the conscious awareness of the imitator,
80 and is termed automatic imitation (Heyes, 2009; 2011). Automatic imitation has been
81 argued to function as a “social glue,” powering cognitive and social development,
82 enhancing emotional reciprocity, and increasing feelings of affiliation, positive rapport
83 and pro-social behaviour (van Baaren et al., 2003; 2009; Lakin & Chartrand, 2003;
84 Chartrand & Bargh, 1999; Cacioppo et al., 2000; Kavanagh & Winkielman, 2016). Given
85 the substantial role automatic imitation plays in our social world, it has received
86 attention from diverse research disciplines. For instance, in social psychology, studies of
87 automatic imitation (termed motor mimicry) have typically involved measuring overt
88 copying behaviours during live social interactions (Chartrand & Bargh, 1999; Ray &
89 Heyes, 2011). In one such study, Chartrand & Bargh (1999) showed that even though it
90 was unrelated to the experimental task, participants copied a confederate’s behaviour
91 such that they touched their face more than waggled their foot when the confederate
92 also touched their face, and vice versa.

93 By contrast, in cognitive psychology, stimulus-response compatibility (SRC)
94 paradigms have been used in order to measure automatic imitation (Brass et al., 2000;
95 Stürmer et al., 2000; Kilner et al., 2003). In a typical SRC task measuring automatic
96 imitation, participants are required to lift their index or middle finger in response to a
97 number cue ('1' for index finger, '2' for middle finger). Simultaneously, they see either
98 the same finger movement (compatible condition) or a different finger movement
99 (incompatible condition). Participants respond slower in the incompatible condition as
100 the observed movement interferes with their response. This difference in reaction time
101 between compatible and incompatible conditions is referred to as the compatibility
102 effect and is considered to be a measure of automatic imitation control. The logic of the
103 task is based on the idea that the observed action generates a motor representation of
104 the same action in the observer. Thus, it has been suggested that on incompatible trials
105 cognitive resources are required to inhibit the automatic tendency to copy an observed
106 (incorrect) action and instead prioritise the alternative (correct) action (Brass & Heyes,
107 2005).

108 After establishing the basic SRC paradigm in order to measure a form of social
109 (imitative) control, subsequent research on automatic imitation in cognitive psychology
110 has provided insight into the factors that influence automatic imitation (Heyes, 2011;
111 Cracco et al., 2018). These studies have typically used an experimental method, which
112 measures the average influence of a manipulation across a group of participants, rather
113 than a differential approach that measures differences across individuals. For example,
114 previous research has found that factors like eye gaze and facial expressions of the
115 interacting partner modulate the tendency to automatically imitate (Wang et al., 2011;
116 Wang & Hamilton, 2014, Grerucci et al., 2013; Crescentini et al., 2011; Rauchbauer et al.,

117 2015, Butler et al., 2016). These findings suggest that social and contextual factors serve
118 as antecedents to automatic imitative behaviours.

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

131 To aid cross-pollination between experimental and differential approaches,
132 more recent imitation research has started to take an individual differences approach
133 by investigating how characteristics of the imitator such as empathy, narcissism,
134 alexithymia and interoceptive awareness, influence automatic imitation (Chartrand &
135 Bargh, 1999; Hogeveen & Obhi, 2013; Obhi et al., 2013; Sowden et al., 2016; Ainley et al.,
136 2014). Such claims, however, are limited due to the small number of studies reported to
137 date, together with the use of relatively small sample sizes and a lack of powerful
138 replications. Moreover, further studies, which used considerably larger sample sizes,
139 have not been able to replicate the moderating influence of personality variables on
140 automatic imitation (Butler et al., 2015; Cracco et al., 2018). Interestingly, however,

141 Butler and colleagues (2015) showed that the sex of the participant modulated the
142 compatibility effect such that females showed a greater compatibility effect compared
143 to males. Therefore, it is possible that biological sex is a factor to consider further when
144 attempting to understand how cognitive mechanisms supporting imitation vary across
145 individuals.

146 Sex is an important individual difference that influences a wide range of
147 cognitive abilities and skills (Geary, 2010), as well as sensitivity to non-verbal social
148 cues (Hall, 1978). However, few studies have investigated how socio-cognitive abilities
149 vary as a function of biological sex and the ones that do have typically focused on
150 mental reasoning or emotion perception (Krach et al., 2009; Russel et al., 2007; Rahman
151 et al., 2004; Campbell et al., 2002). Further, such prior studies have often produced
152 mixed results based on relatively small sample sizes (Hyde, 2014; Miller & Halpern,
153 2014). Therefore, the potential influence of sex on complex cognitive mechanisms that
154 control non-verbal interactions, remains largely unknown.

155 The extent to which sex differences operate in imitative behaviour has also
156 received minimal attention to date. For example, no sex differences have been found on
157 the automatic imitation of actions or gestures (Chartrand & Bargh, 1999; Larsen et al.,
158 2010). By contrast, studies on facial mimicry have shown that females automatically
159 imitate facial expressions more than males (Dimberg, 1990; Sonnyby-Borgstrom et al.,
160 2008). Although there is no consistent empirical evidence to date, which speaks to sex
161 differences in imitation, there is theoretical reason to think that sex differences may
162 exist in imitative behaviour. Indeed, there is robust evidence for females to be more
163 empathetic than males (Christov-Moore et al., 2014; Baron-Cohen & Wheelwright,
164 2004). Further, empathy has been associated with a variety of paradigms investigating

165 imitation (Mueller et al., 2013; Sonnyby-Borgstrom 2002; Chartrand & Bargh, 1999).
166 Therefore, given that females are likely to be more empathetic than males in general,
167 this may lead to more imitative tendencies in particular social contexts.

168 A core question pertains to whether the sex difference seen in the SRC task is a
169 genuine difference between males and females or reflects an in-group/own-sex bias.
170 The stimuli used in the task by Butler and colleagues (2015) were of a female hand. It
171 may be that automatic imitation increases when the sex of the participant and the
172 interacting partner are matched. Group biases (typically in-group favouritism and out-
173 group dislike) are prevalent in day-to-day social interactions (Allport, 1954; Cameron et
174 al., 2001), relating to race, ethnicity (Ito & Bartholow, 2009; Kubota et al., 2012; Malpass
175 & Kravitz, 1969; van Bavel & Cunningham, 2009; Milner, 1983; Aboud 1988;), sex
176 (Brown, 1995; Yee & Brown, 1994; Fishbein, 1996, Powlishta, 1995; Rudman &
177 Goodwin, 2004), and arbitrary groups (Tajfel et al., 1971; Bernstein et al., 2007).
178 Therefore, ingroup biases seem like a powerful mechanism, which may guide imitative
179 behaviour based on the sex of the interaction partner.

180 In imitation research specifically, children have been known to imitate same-
181 sex models more than others (Shutts et al., 2010). Facial imitation and SRC measures of
182 automatic imitation have both been found to increase when the interacting partner is an
183 in-group member compared to an out-group member based on race, ethnicity and
184 arbitrary group assignment (Mondillon et al., 2007; Rauchbauer et al., 2015; Gleibs et
185 al., 2016). Moreover, recent work provides suggestive evidence for a sex difference
186 and/or in-group bias in the automatic imitation task (Cracco et al., 2018; Genschow et
187 al., 2017). For example, a meta-analysis found a higher reaction time compatibility
188 effect when the sex of the stimuli matched the sex of the majority of participants in the

189 sample (Cracco et al., 2018)¹. However, the extent to which this sex difference reflects
190 an in-group bias remains unclear because no existing study has manipulated the sex of
191 the stimuli across male and female participants.

192 Two other possible explanations exist. First, the sex difference on the automatic
193 imitation SRC task could reflect that females tend to automatically imitate more than
194 males, and therefore require more cognitive resources to inhibit the tendency to
195 automatically imitate, leading to a greater compatibility effect. If so, the sex difference
196 would be tied to a process related to imitation specifically. Second, the sex difference
197 may be more domain-general in nature i.e. it may reflect a basic difference in the
198 cognitive systems that underlie performance on SRC tasks more generally. Consistent
199 with a domain-general explanation, sex differences have been found on many non-social
200 inhibitory control tasks which, like the imitation task, require the inhibition of task-
201 irrelevant automatic response tendencies in order to enforce a task-relevant response
202 (e.g. flanker, oddball, gaze- and arrow-cueing, and Simon tasks; Stoet, 2010; Judge &
203 Taylor, 2012; Clayson et al., 2011; Rubia et al., 2010; Bayliss et al., 2005; Merritt et al.,
204 2007; Alwall et al., 2010). As such, sex differences in SRC tasks may reflect differences in
205 cognitive systems that operate across these tasks such as selective attention (Clayson et
206 al., 2011) and/or spatial processing (Stoet, 2017).

207 These findings suggest that it is as yet unclear whether the sex difference on
208 SRC measures of automatic imitation reflect more domain-general processes or
209 processes solely tied to imitative control (Butler et al., 2015; Cracco et al., 2018;

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.

210 Genschow et al., 2017; Darda et al., 2018; Darda & Ramsey, 2019). The SRC task used by
211 researchers to demonstrate the existence of a sex difference was a composite of both
212 spatial and imitative components (Butler et al., 2015; Genschow et al., 2017). A sex
213 difference solely tied to imitative control might suggest that a distinct mechanism, or a
214 partially distinct set of mechanisms, may underpin performance on the automatic
215 imitation task compared to other inhibitory control tasks. Thus, in order to understand
216 the cognitive architecture of social interactions, it is critical to unpack the relative
217 contributions of both general and specific components in socio-cognitive processes
218 (Michael & D’Ausilio, 2015; Spunt & Adolphs, 2017; Binney & Ramsey, in 2020; Ramsey,
219 2018; Ramsey & Ward, 2020). Therefore, in the current study, we investigate sex
220 differences on the automatic imitation task as well as a non-social control task in order
221 to investigate whether the sex difference relies on domain-general and/or domain-
222 specific mechanisms.

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

235 underlying the sex difference are domain-general or domain-specific (or a combination
236 of both).

237 **Experiment 1**

238 **Introduction**

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

247 Further, we also investigate the extent to which stable dimensions of
248 personality influence the control of automatic imitation as measured on the SRC task.
249 Prior work has provided mixed evidence regarding this question. Some studies have
250 found a link between automatic imitation and empathy and narcissism - automatic
251 imitation was higher for individuals who scored high on the empathy scale, and lower
252 for those who scored higher on the narcissism scale. (Chartrand & Bargh, 1999; Obhi et
253 al., 2013; Hogeveen & Obhi, 2013). There are theoretical grounds to also posit a link
254 between automatic imitation and two of the Big Five personality factors. Agreeableness
255 and extraversion have been previously linked to empathy, altruism, and sociability
256 (Barrio et al., 2004; Ashton et al., 1998; McCrae & Costa, 1999), and are thus considered
257 as contributors to prosocial behaviour (Graziano & Eisenberg, 1997). Thus, individuals
258 who are more agreeable and more extraverted may be more prosocial and could thus

259 imitate their interacting partners more than others. In addition, although debated
260 (Southgate & Hamilton, 2008; Hamilton, 2013), imitation abilities have been argued to
261 vary in atypical populations including autism spectrum disorders and schizophrenia
262 (Oberman & Ramachandran, 2007; Thakkar et al., 2014; Williams et al., 2001),
263 indicating that a relationship may exist between autistic-like and schizotypal traits and
264 automatic imitation.

265 The largest datasets to date, however, show that performance on the SRC task is
266 invariant to stable personality variables (Butler et al., 2015; Cracco et al., 2018). One
267 concern with such null effects of personality is that they may reflect the impoverished
268 social context of the SRC task. That is, effects of interest may only operate in more
269 socially meaningful contexts. Therefore, in Experiment 1, we make the social context
270 more meaningful by including emotional facial expressions within our design and
271 investigate the extent to which automatic imitation continues to remain invariant as a
272 function of personality. We included five face images depicting five emotional
273 expressions (fearful, angry, happy, sad, neutral), and personality variables included
274 extraversion, agreeableness, autistic-like and schizotypal traits, narcissism (grandiose
275 and vulnerability narcissism), empathy (empathic concern and perspective taking), and
276 alexithymia (for detailed information about measures used, see supplementary
277 material). We included only the perspective taking and empathic concern subscales of
278 the empathy questionnaire as we had directional predictions about these subscales –
279 prior evidence suggests a positive link between perspective taking and imitation, as
280 well as empathic concern and imitation (Santesteban et al., 2012; Chartrand & Bargh,
281 1999). Following Butler and colleagues (2015), although there is reason to expect pro-
282 social dimensions of personality to be related to increased imitation, based on prior

283 empirical evidence using the SRC task, we would expect imitation to be invariant to
 284 stable dimensions of personality.

285 **Method**

286 Across all experiments, we report how the sample size was determined, all data
 287 exclusions, and all measures in the study (Simmons et al., 2011; 2012). Following open
 288 science initiatives, all raw data are available online for other researchers to pursue
 289 alternative questions of interest. For all three experiments, data pre-processing,
 290 statistical analyses, and data visualisations were performed using R (R Core Team,
 291 2018), unless otherwise specified. All raw data and code used for analyses are available
 292 online (<https://osf.io/fsh9b/>). For all following experiments, we report our primary
 293 and secondary questions of interest in Table 1.

294

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?

295

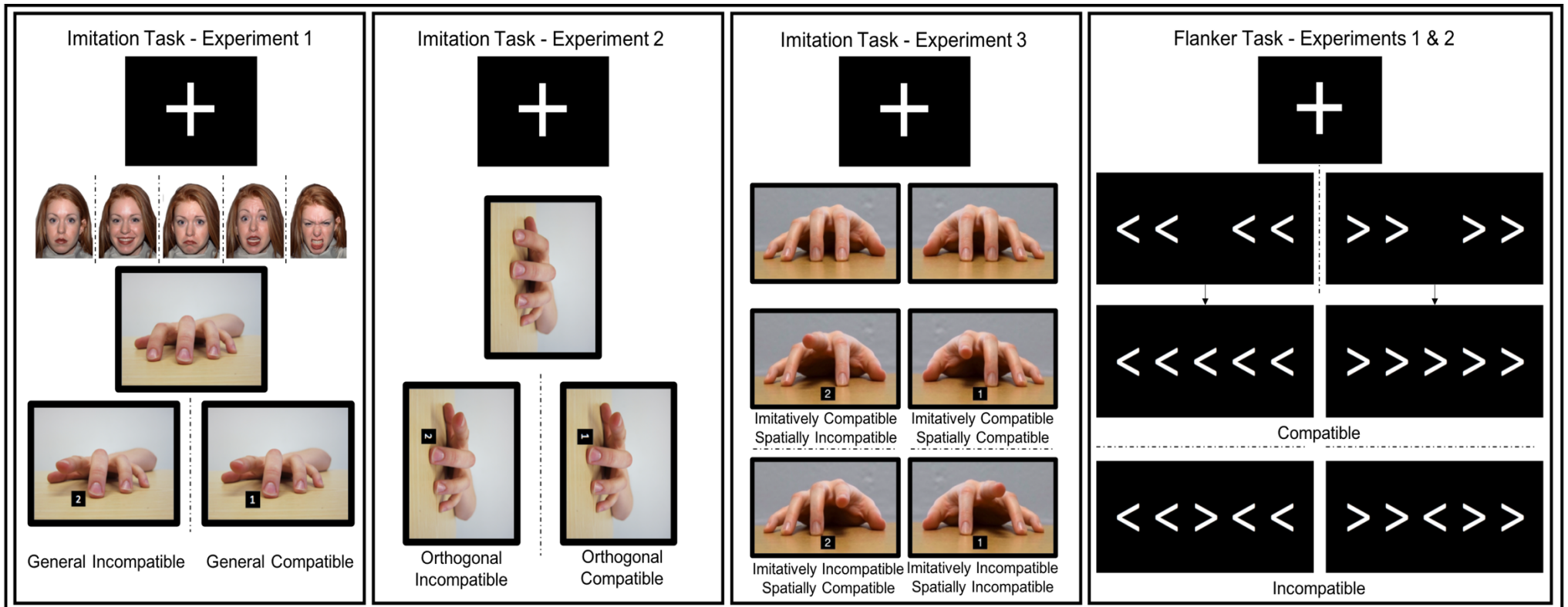
296 *Table 1.* Table 1 highlights primary and secondary questions of interest for experiments 1, 2,
297 and 3 of this paper.

298

299 We determined the sample size for our experiments as follows. For experiment
300 1, we aimed to collect as many participants as possible over a two-day data collecting
301 session. Therefore, the stopping rule was to terminate data collection after day 2 of data
302 collection. For Experiments 2 and 3, in order to focus our design on the primary
303 research question, which concerned sex differences, we set a minimum sample size of
304 100 male and 100 female participants. Sensitivity analyses revealed that given a sample
305 size of two hundred participants (100 per sex), we would have 80% power to detect an
306 effect size of Cohen's $d > 0.35$ for the mean difference between the two sexes on a one-
307 tailed t-test, and an effect size of $\eta_p^2 > 0.04$ for a 2 x 2 mixed ANOVA. Such a design,
308 therefore, provides reasonable confidence (80%) to detect effect sizes of interest that
309 are conventionally considered small-to-medium.

310 *Participants*

311 Two hundred and three participants took part in this experiment for monetary
312 compensation (£6) or course credit. All participants provided informed consent, had
313 normal or corrected-to-normal vision, and were right-handed. Approval was obtained
314 from the Research Ethics and Governance Committee of the School of Psychology at
315 Bangor University. Participants were excluded if performance was three standard
316 deviations away from the group mean average performance per condition in terms of
317 accuracy or reaction time (N=14 for the imitation task, N=7 for the flanker task). A
318 further 14 participants were excluded as demographic information (sex of the
319 participant, or both age and sex of the participant) was not recorded. For the imitation
320 task, the final sample included 175 participants (59 males, $Mean_{age} = 20.9$, $SD_{age} = 4.23$).
321 For the flanker task, the final sample included 182 participants (59 males, $Mean_{age} =$
322 20.9 , $SD_{age} = 3.73$).



3
4
5 **Figure 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
6 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
7 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
8 to measure imitative and spatial effects independent of each other.

330 *Stimuli, tasks, and procedure*

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

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

353 Each trial began with the presentation of a fixation cross for 500 milliseconds
354 (ms). A face image was presented at the top of the screen after the fixation cross for 500
355 ms, followed by the neutral hand image in the centre of the screen. The face image
356 remained on the screen above the neutral hand and target hand image for the
357 remainder of the trial. The neutral hand was presented for a random inter-stimulus
358 interval (ISI) of 500, 700, or 1000 ms, followed by the target hand image. The
359 succession of neutral and target hand images was such that it produced apparent
360 motion of either an index or middle finger lift simultaneously with the presentation of
361 the number cue. The target hand image remained on the screen until the participant
362 made a response (but no longer than 2000 ms). The total trial length varied depending
363 on the ISI, but was never longer than 3500 ms. Trials were pseudo-randomised in such a
364 way that no more than 4 identical trials were presented consecutively. There were four
365 blocks of 50 trials each which included 25 compatible trials and 25 incompatible trials
366 with equal number of trials per face image.

367 *Flanker task.* The flanker task was based on the paradigm developed by Erikson
368 and Erikson (1974; Figure 1). The stimuli consisted of five equally sized and spaced
369 white arrows on a black background. Participants were instructed to respond to the
370 direction of the central arrow – they were asked to press key ‘m’ with their right index
371 finger if the central arrow pointed to the right, and press key ‘n’ with their left index
372 finger if the central arrow pointed to the left. The direction of the flanker arrows was
373 either compatible (<<<<< OR >>>>>) or incompatible (<<><< OR >><>>) to the central
374 arrow direction. This produced four trial types and two conditions (compatible and
375 incompatible).

376 Each trial started with a fixation cross for 800 ms, 1000 ms, or 1200 ms. The
377 flanker arrows then appeared on the screen for 100 ms, followed by the central arrow

378 in between the flankers. The five arrows remained on the screen until the participant
379 responded (but for no longer than 1600 ms). Participants were first presented with the
380 fixation cross for 800 ms, 1000 ms, or 1200 ms, followed by the presentation of the four
381 flanker arrows for 100 ms. Total trial length was never longer than 2900 ms. Trials
382 were pseudo-randomised in such a way that no more than 4 identical trials were
383 presented consecutively. Each participant did one block of 64 trials, with 32 compatible,
384 and 32 incompatible trials. Further, in this experiment, we addressed an additional
385 unrelated question – in half of the compatible and incompatible trials, flanker arrows
386 flipped arrow direction during the trial between their initial presentation on the screen
387 and the appearance of the central arrow. However, as we were interested in the basic
388 compatibility effect, we collapsed trials across conditions irrespective of whether they
389 changed direction mid trial or not.

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

392 *Questionnaires.* Participants also completed a range of self-report questionnaires
393 which included the Mini International Personality Item Pool (mini IPIP; Donnellan et al.,
394 2005; the Short Autism Spectrum Quotient (AQ-10 Adult; Baron-Cohen et al., 2001;
395 Allison et al., 2012), the Brief Schizotypal Personality Questionnaire (SPQ-B; Raine &
396 Benishay, 2005), the Narcissistic Personality Inventory (NPI-16; Ames et al., 2006), the
397 Hypersensitivity Narcissism Scale (HSNS; Hendin & Cheek, 1997), the Interpersonal
398 Reactivity Index (IRI; Davis, 1980), and the Toronto Alexithymia Scale (TAS-20; Bagby
399 et al., 1994). For more details on the measures used and how the questionnaires were
400 scored, see the Supplementary Material.

401 *Data Analysis*

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

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

417 Data was analysed as follows: first, for both the RT and accuracy data on the
418 imitation task, a 2 (compatibility: incompatible, compatible) x 5 (emotion: neutral, sad,
419 happy, fearful, angry) repeated measures ANOVA was performed to investigate whether
420 facial cues signalling emotional states modulated the compatibility effect on the
421 imitation task. Second, on both RT and accuracy data, for the flanker and imitation tasks
422 separately, a 2 (compatibility: compatible, incompatible) x 2 (sex: male, female) mixed
423 ANOVA was performed in order to investigate whether the compatibility effect on the
424 imitation and flanker tasks varies as a function of sex. For all ANOVAs, we report
425 Greenhouse Geisser corrected values if the assumption of sphericity is violated. Third,
426 in order to investigate whether the flanker and imitation compatibility effects were

427 correlated, a one-tailed Pearson's correlation was performed. A positive correlation
428 would suggest that the two compatibility effects were related to each other.

429 Based on prior research (Heyes, 2011; Brass et al., 2000; Eriksen & Eriksen,
430 1974), we expected a main effect of compatibility on the 2 x 2 ANOVA such that RT
431 would be higher, and accuracy would be lower on incompatible trials compared to
432 compatible trials. In support of our hypothesis, we also expected a Compatibility*Sex
433 interaction such that the compatibility effect would be higher for females as compared
434 to males. The interaction effect was central to testing our primary hypothesis, and thus,
435 we calculated compatibility effects for male and female participants separately by
436 computing the mean difference and 95% confidence intervals between compatible and
437 incompatible conditions. In order to directly estimate the size of the difference in
438 compatibility effects between males and females, we then again computed the mean
439 difference and 95% confidence interval. We used one-tailed 95% confidence intervals
440 as we had a directional hypothesis that females would have a higher compatibility effect
441 than males on both the imitation and flanker tasks (Butler et al., 2015; Stoet, 2011;
442 Clayson et al., 2011). Further, we plot the group-average data and individual participant
443 data in the results section to ensure that any effects that go in a direction contrary to
444 our hypothesis can be visualised. We also provide all raw data online so that
445 researchers can test for alternative hypotheses.

446 We also report standardised effect sizes for ANOVA using partial eta-squared (η)
447 for independent samples t-tests using Cohen's d and for paired samples t-tests using
448 Cohen's d_z (Cohen, 1992; Lakens, 2013). We also report and interpret the point and
449 interval estimate using 95% CIs for effect sizes of interest in line with recent
450 suggestions (Cumming, 2012; Amrhein et al., 2019). In order to quantify the evidence
451 for a null hypothesis over the experimental hypothesis (where a null result was found

452 using null hypothesis significance testing; NHST), we calculated the Bayes factor (BF_{01})
453 by performing a Bayesian independent samples t-test to investigate the sex difference
454 between the sexes. We used default priors in JASP for the independent samples t-test,
455 that is, a Cauchy distribution³ with spread r set to $1/\sqrt{2}$. The Bayes factor was
456 interpreted using benchmark criteria from Jeffereys (1961). Bayesian analyses, Cohen's
457 d and d_z , as well as 95% CIs were calculated using JASP (JASP Team, 2018).

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

477 the SRC imitation task. As sex differences have been previously found on personality
478 measures (Schmitt et al., 2008), we computed sex*trait interaction terms for all
479 personality variables, and evaluated them in separate multiple regression models.

480 For all questionnaires used, we report reliability information as indexed by
481 Cronbach's alpha (Cronbach, 1951). For the imitation and flanker task compatibility
482 effects, we report the internal consistency of the measure by calculating split-half
483 reliability using a permutation-based split-half approach with 5000 random splits
484 (Parsons, 2018) using the splithalf package in R. All reliability analyses are reported in
485 the Supplementary Material.

486 Throughout the paper, following Gigerenzer (2018), we avoid interpreting
487 results based solely on p values and a binary distinction between "significant" and "non-
488 significant". Instead, we base the direction and strength of our interpretation on a range
489 of metrics, which include a p value and an associated measure of sensitivity (power),
490 effect sizes in original and standardised units along with a measure of precision using
491 95% confidence intervals (Cumming, 2012). Further, we run multiple replication and
492 extension experiments (Zwaan et al., 2018) and we meta-analyse the main effects
493 across experiments (Cumming, 2012), both of which help to further aid the confidence
494 that we can have in our findings and the credibility of the conclusions more generally.

495 **Results**

496 *Automatic imitation task*

497 *Accuracy.* Average accuracy on the imitation task was above 90% for both males and
498 females on both compatible and incompatible conditions (Supplementary Figure 1,
499 Supplementary Table 1). A 2 (compatibility: compatible, incompatible) x 5 (emotion:
500 neutral, sad, happy, fearful, angry) ANOVA showed no main effect of emotion ($F(4, 696)$)

501 = 0.50, $p=0.729$, $\eta_p^2 = 0.003$) and no Compatibility*Emotion interaction ($F(4, 656) =$
502 1.20 , $p=0.31$, $\eta_p^2 = 0.007$). Thus, for all further analyses of accuracy, trials are collapsed
503 across all emotion conditions.

504 **Table 2.**

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

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

506

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

511

512

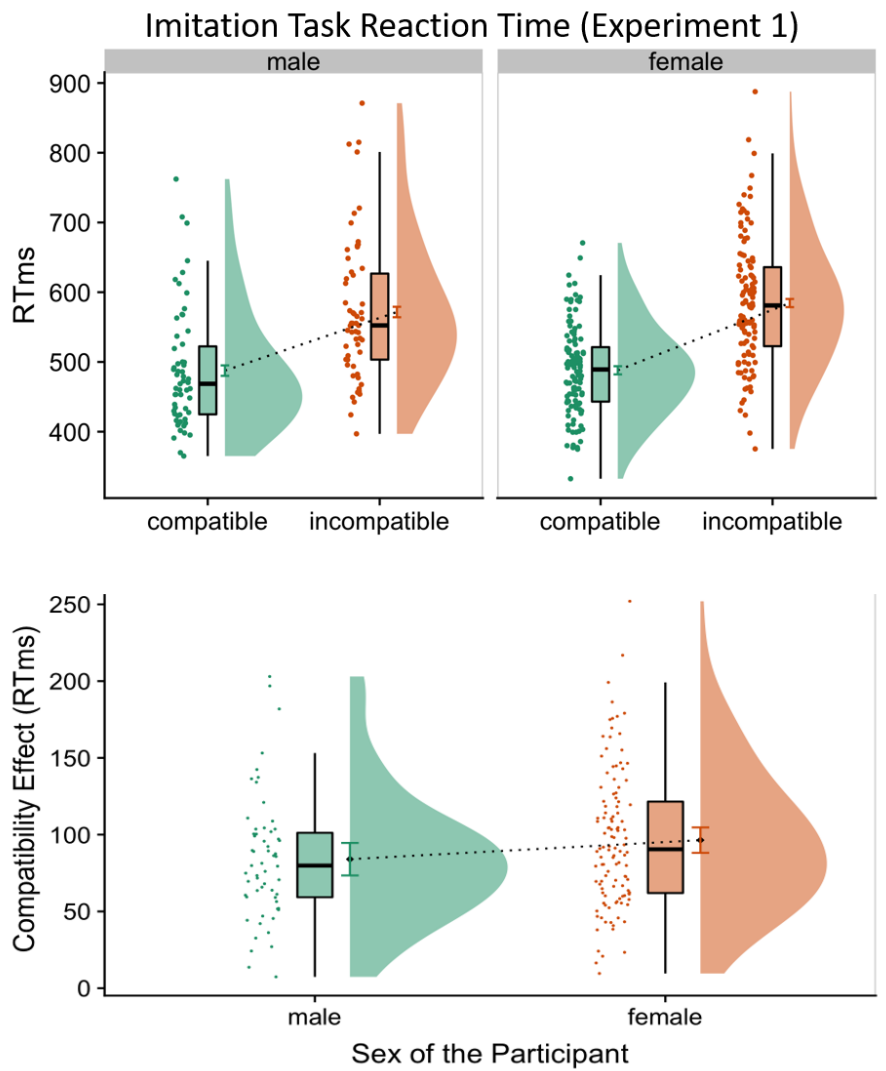
513 The 2 (compatibility: compatible, incompatible) x 2 (sex: male, female) mixed
514 ANOVA showed a main effect of compatibility such that participants were more
515 accurate on compatible trials than incompatible trials ($F(1, 173) = 258.09, p < .001, \eta_p^2 =$
516 0.60 ; Supplementary Figure 1). The effect size for the main effect of compatibility is
517 conventionally considered to be large. For the main effect of sex ($F(1, 173) = 0.22, p =$
518 $0.64, \eta_p^2 = 0.001$) and the Compatibility*Sex interaction ($F(1, 173) = 0.60, p = 0.44, \eta_p^2 =$
519 0.003), the effect sizes were close to zero (Supplementary Table 2).

520 *Reaction time.* Average reaction times on the imitation task for both males and
521 females on both compatible and incompatible conditions were between 485 and 585
522 milliseconds (Figure 2, Supplementary Table 1). A 2 (compatibility: compatible,
523 incompatible) x 5 (emotion: neutral, sad, happy, fearful, angry) ANOVA showed no main
524 effect of emotion ($F(4, 696) = 1.81, p = 0.127, \eta_p^2 = 0.004$). Importantly, the effect size for
525 the Compatibility*Emotion interaction was close to zero ($F(4, 696) = 0.40, p = 0.796, \eta_p^2$
526 $= 0.002$, see Supplementary Figure 2). Thus, for all further analyses of RT, trials are
527 collapsed across all emotion conditions.

528 The 2 (compatibility: compatible, incompatible) x 2 (sex: male, female) mixed
529 ANOVA showed a main effect of compatibility such that participants were slower to

530 respond on incompatible trials than compatible trials ($F(1, 173) = 669.77, p < .001, \eta_p^2 =$
531 0.80 ; Figure 2). The effect size for the main effect of compatibility is conventionally
532 considered to be large. Effect size for the main effect of sex was close to zero with a p-
533 value of $p = 0.61$ ($F(1, 173) = 0.26, p = 0.61, \eta_p^2 = 0.001$). The Compatibility*Sex
534 interaction had a small effect size ($F(1, 173) = 3.16, p = 0.08, \eta_p^2 = 0.018$; Supplementary
535 Table 2).

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



551
 552 **Figure 2.** *Experiment 1 – Imitation Task Reaction Time.* Reaction time is reported in milliseconds
 553 (ms). The upper panel shows mean reaction times for compatible and incompatible conditions for
 554 both males and females. The lower panel shows the compatibility effect for both males and females.
 555 The compatibility effect is calculated by subtracting reaction times on compatible trials from
 556 incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms = reaction
 557 time in milliseconds.

558

559 *Flanker task*

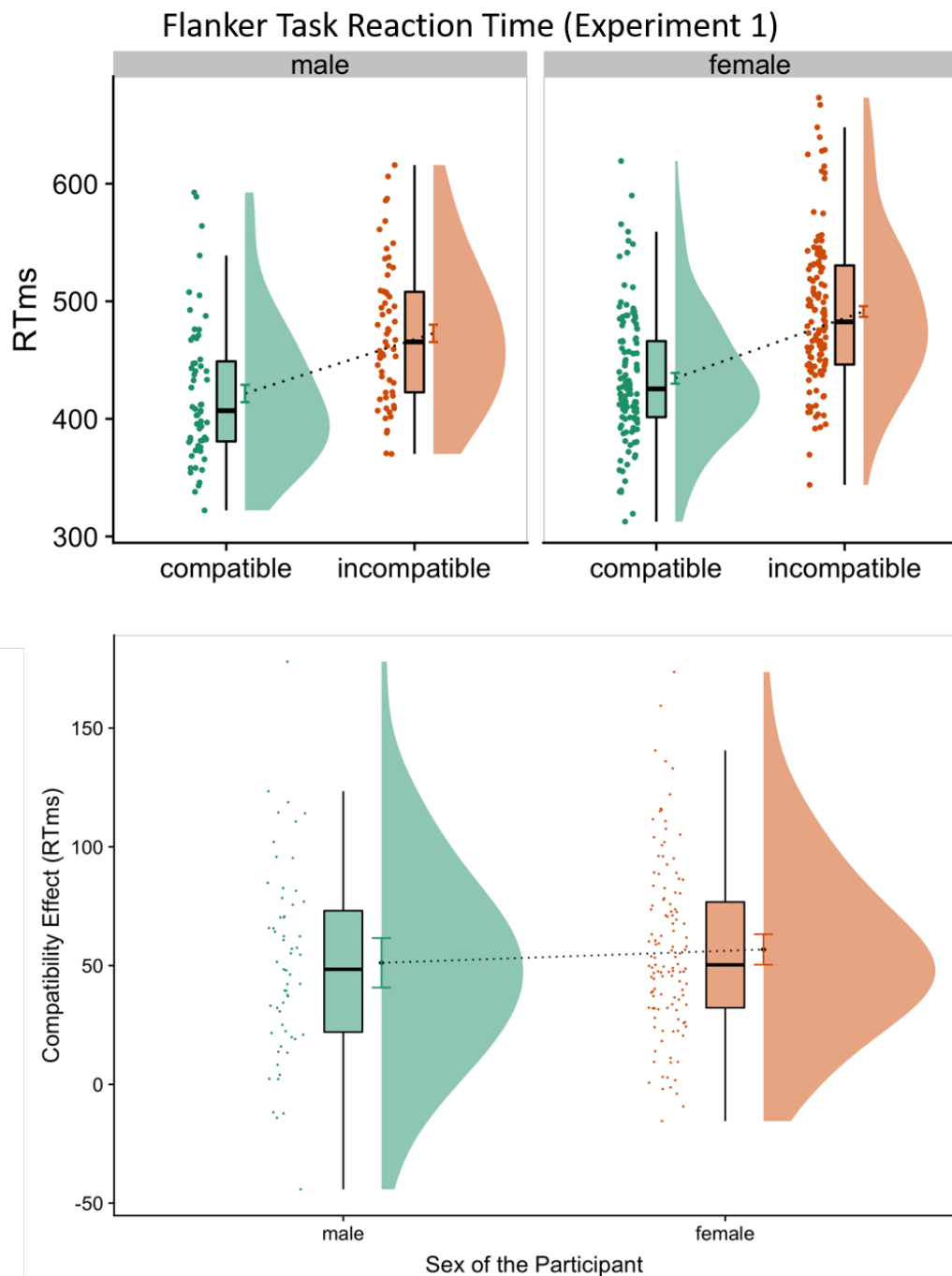
560 *Accuracy.* Average accuracy on the flanker task was above 94% for both males
 561 and females on both compatible and incompatible conditions (Supplementary Figure 3,
 562 Supplementary Table 1). A 2 (compatibility: compatible, incompatible) x 2 (sex: male,
 563 female) mixed ANOVA was performed. The effect size for the main effect of
 564 compatibility was close to zero ($F(1, 180) = 2.24, p = 0.136, \eta_p^2 = 0.01$). Effect sizes for
 565 the main effect of sex and ($F(1, 180) = 0.04, p = 0.85, \eta_p^2 = <0.001$) and the

566 Compatibility*Sex interaction ($F(1, 180) = 0.09, p = 0.759, \eta_p^2 < 0.001$) were also close to
567 zero with p-values > 0.1 (Supplementary Table 2).

568 *Reaction time.* Mean reaction time on the flanker task for both males and females
569 on both compatible and incompatible conditions was between 420 and 495
570 milliseconds (Figure 3, Supplementary Table 1). A 2 (compatibility: compatible,
571 incompatible) x 2 (sex: male, female) mixed ANOVA (Figure 3) showed a main effect of
572 compatibility such that participants were slower to respond on incompatible trials than
573 compatible trials ($F(1, 180) = 334.15, p < .001, \eta_p^2 = 0.65$). The effect size for the main
574 effect of compatibility is conventionally considered to be large. The main effect of sex
575 had a small effect size and showed that overall females were slower than males on the
576 flanker task ($F(1, 180) = 3.40, p = 0.08, \eta_p^2 = 0.02$). The Compatibility*Sex interaction
577 had an effect size close to zero with a p-value of 0.34 ($F(1, 180) = 0.90, p = 0.34, \eta_p^2 =$
578 0.005 ; see Supplementary Table 2).

579 To further compare with the automatic imitation task, compatibility effects in the
580 flanker task were computed separately for males and females, and then compared to
581 each other. For both males and females, compatibility effects had a large effect size
582 (Cohen's $d_z > 1.2$) with the lower bound of the 95% confidence interval at 0.98 or
583 higher. When compatibility effects for males and females were directly compared to
584 each other, there was a trend toward females showing a higher compatibility effect than
585 males by 5.62ms, with the lower bound of the 95% confidence interval at -4.14 ms
586 below zero (Mean Difference = 5.62 ms, 95% CI[-4.14, ∞], Cohen's $d = 0.15, 95\% CI [-$
587 $0.11]$; see Figure 3, Table 2A). The effect size was small, with the lower bound of the
588 95% confidence interval at -0.11. Thus, a reasonable estimate for the mean difference
589 between males and females on the flanker compatibility effect ranges from -4.14ms to

590 5.62 ms, with one possibility being a small difference between females and males, such
591 that females may show a higher compatibility effect than males. However, a Bayesian
592 independent samples t-test showed that the data was 3 to 4 times more likely under the
593 null hypothesis than the alternative hypothesis ($BF_{01} = 3.85$).



594 **Figure 3.** Experiment 1 – Flanker Task Reaction Time. Reaction time is reported in milliseconds
595 (ms). The upper panel shows mean reaction times for compatible and incompatible conditions
596 for both males and females. The lower panel shows the compatibility effect for both males and
597 females. The compatibility effect is calculated by subtracting reaction times on compatible trials
598

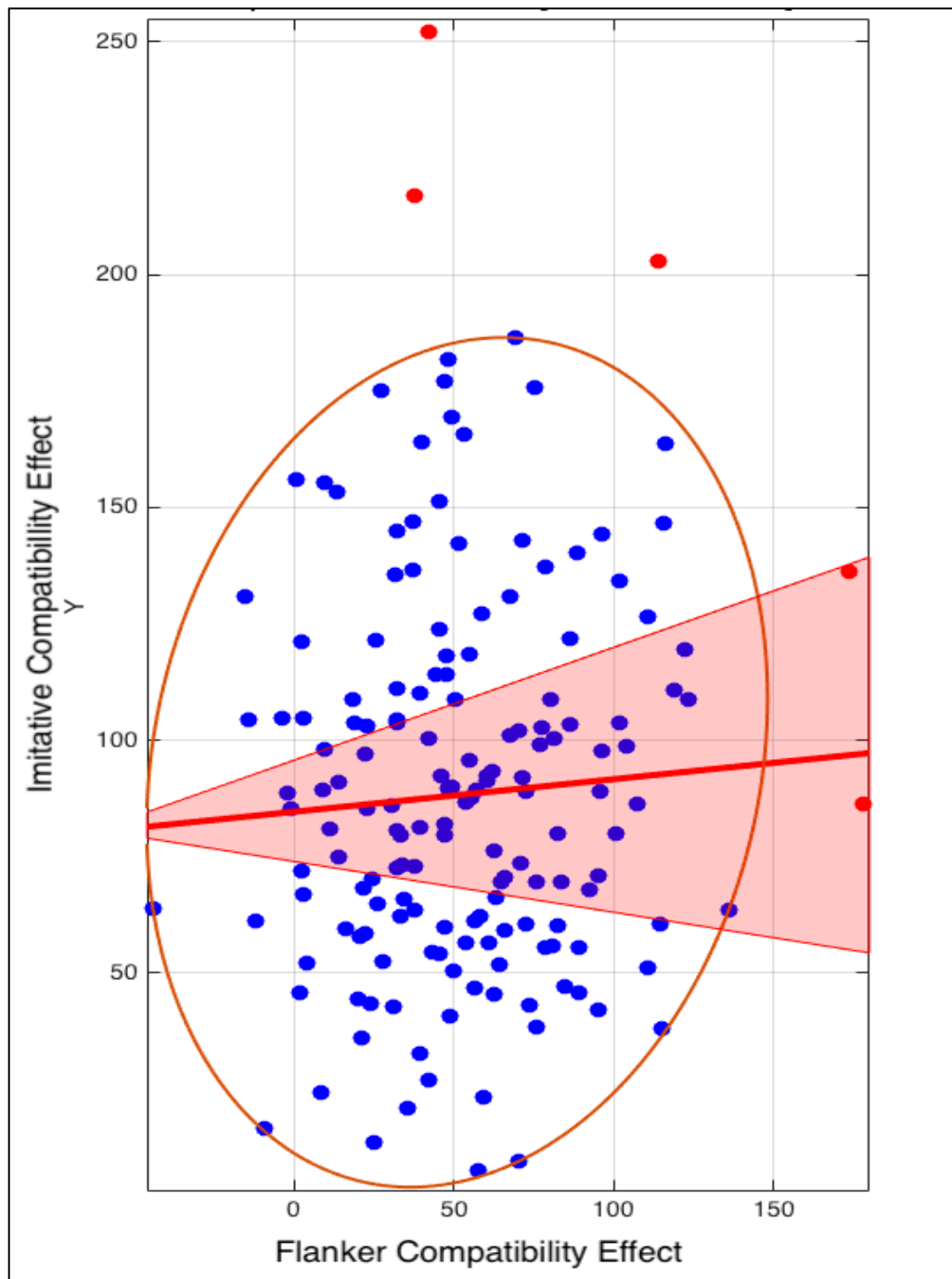
599 from incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms =
600 reaction time in milliseconds.

601

602 *Correlational analysis*

603 In order to investigate whether the flanker and imitation compatibility effects
604 were correlated, a one-tailed skipped correlation was performed. To do so, only those
605 participants who performed both the tasks were included in the analysis (N=165). The
606 skipped correlation analyses were performed using a Matlab-based toolbox
607 (Mathworks Inc., MA; <http://sourceforge.net/projects/robustcorrtool/>, Pernet, Wilcox,
608 & Rousselet, 2013). Skipped correlation takes into consideration the overall structure of
609 the data, and protects against bivariate outliers. In order to perform a skipped
610 correlation analysis, we first tested the assumptions of normality and homogeneity of
611 variance. The Henze-Zirkler's multivariate normality test (Trujillo-Ortiz et al., 2007)
612 indicated that the data was close to normally distributed, and the test for heterogeneity
613 indicated that the data have the same variance. Next, we estimated the robust centre of
614 the data using the minimum covariance determinant (MCD) estimator. The MCD
615 estimator is considered to be a robust estimator of the scatter and location of
616 multivariate data (Rousseeuw, 1984; Rousseeuw & van Drissen, 1999; Verboten &
617 Hubert, 2005). Bivariate outliers were then identified by using a projection technique –
618 data points were orthogonally projected by lines joining each data point to the robust
619 centre of the data cloud. Five bivariate outliers were removed using the box-plot rule
620 relying on the interquartile range (Carling, 2000), and skipped correlation was
621 computed on the remaining data. Following guidelines put forward previously (Pernet
622 et al., 2013), as the data were close to being normally distributed, we used a skipped
623 Spearman correlation analysis. A one-tailed skipped Spearman correlation analysis
624 showed a small positive correlation between the imitation and flanker compatibility

625 effects, which did not pass our statistical threshold, with the lower bound of the 95% CI
626 extended below zero ($r(160) = 0.04$, 95% CI $[-0.12, \infty]$; see Figure 4).



627
628 **Figure 4.** *Experiment 1 - Correlation Analysis.* A skipped Spearman correlation shows a small
629 positive correlation between the flanker and imitative compatibility effects that does not pass
630 our statistical thresholding. Abbreviations: RTms (reaction time in milliseconds). Dots in red are
631 the bivariate outliers.

632

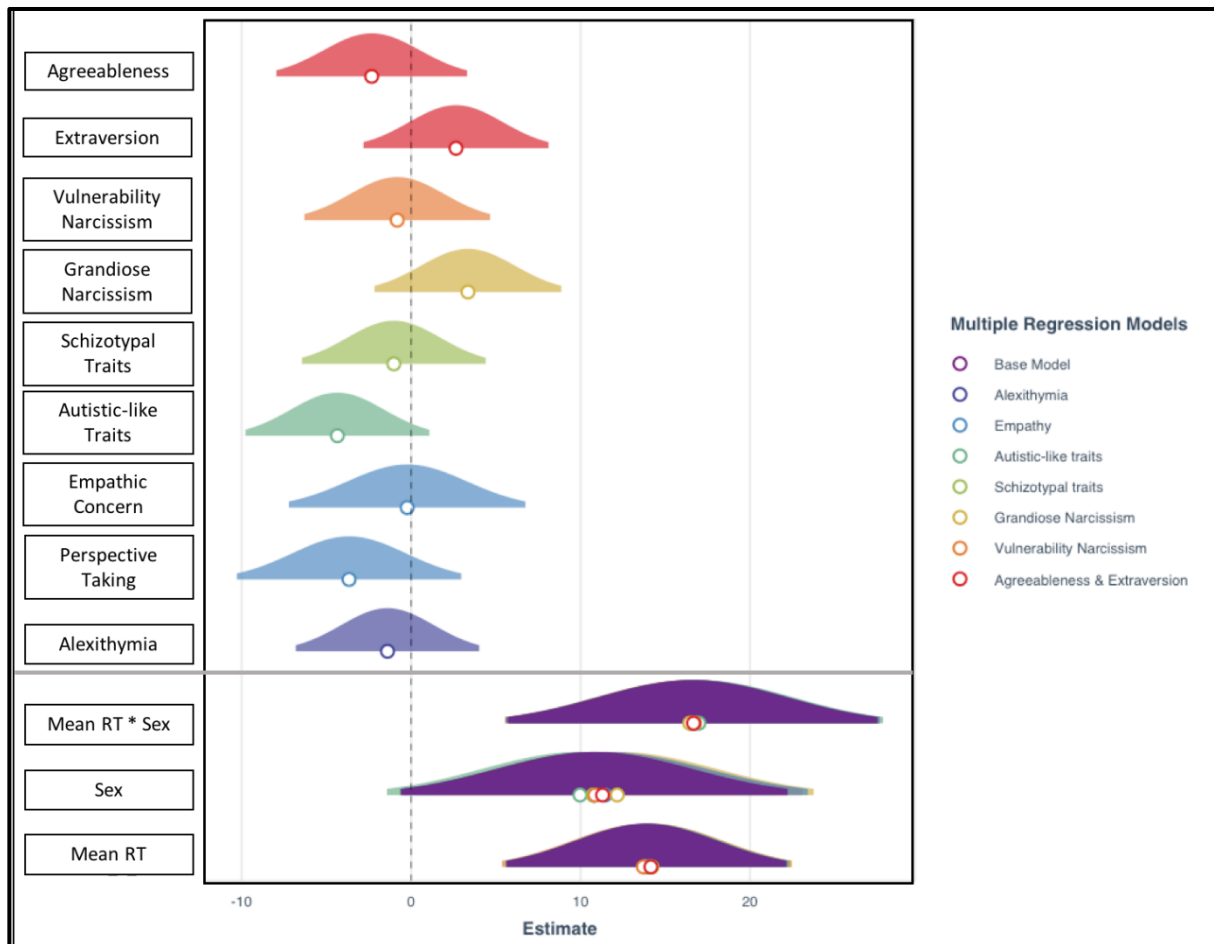
633 Our results thus suggest that the compatibility effects across these tasks were
634 largely unrelated, and participants with greater interference in one task did not
635 experience a greater interference in the other task.

636 *Multiple regression analyses*

637 We also investigated the relationship between stable personality measures and
638 the general compatibility effect as measured on the SRC task. Tests for multicollinearity
639 indicated that a very low level of multicollinearity was present (*VIF* for all predictor
640 variables < 2). The base model included mean RT, sex, and a meanRT*sex interaction
641 term. The base model explained 33.6 % of the variance in the congruency effect
642 ($F(3,171)=28.88$, $p<.001$, $R^2=.34$, $f^2=0.51$) and indicated a medium effect size (Cohen,
643 1992). Mean RT predicted the compatibility effect, with increasing CE as mean RT
644 increases ($B=0.27$, $SEB=0.03$, $t(171)=8.02$, $p<.001$, 95% CI [0.20; 0.34]). In addition to
645 mean RT, sex marginally predicted the compatibility effect ($B=5.40$, $SEB=2.88$,
646 $t(171)=1.87$, $p=.06$, [-0.29; 11.10]) with a higher compatibility effect for females than
647 males. The mean RT * sex interaction was also a predictor ($B=0.10$, $SEB=0.03$,
648 $t(171)=3.01$, $p=.003$, [0.03; 0.17]), suggesting that increases in mean RT predicted
649 larger increases in the compatibility effect for females ($B=0.37$, $SEB=0.04$, $t(171)=8.43$,
650 $p=.001$) compared to males ($B=0.17$, $SEB=0.05$, $t(171)=3.32$, $p=.001$). Results from the
651 base model are very similar to the results of prior work using a same SRC task and
652 analytical approach (Butler et al., 2015).

653 Agreeableness, extraversion, grandiose and vulnerability narcissism, empathy,
654 autistic-like and schizotypal traits, and alexithymia did not predict the general
655 compatibility effect above and beyond the base model (all p 's > .11, all CIs overlapping
656 with zero; see Figure 5). Effect sizes attributable to the addition of the personality

657 variables (beyond the base model) indicated very small effects (Cohen's f^2 for all models
 658 <0.01 ; Cohen, 1992). The multiple regression models are summarized in Supplementary
 659 Table 3. Zero-order correlations are also consistent with the findings from the multiple
 660 regression analyses, such that there are no relationships between stable personality
 661 measures and CE (see Supplementary Table 4, Supplementary Figure 4).



662
 663 **Figure 5.** *Experiment 1 – Multiple Regression Analyses.* Values of standardised coefficients are
 664 plotted for each predictor variable (personality trait) along with their corresponding
 665 uncertainties (95% confidence interval width for a normal distribution for each estimate).
 666 Coefficients are standardised by dividing by two standard deviation units according to Gelman
 667 (2008). The base model consists in the bottom three predictor variables (depicted in violet) –
 668 mean RT, Sex, and meanRT*Sex. Abbreviations: RT = Reaction Time. *N.B.* The circles on the
 669 purple distributions represent the standardised co-efficients for the main effect of meanRT,
 670 main effect of sex, and meanRT*sex interaction respectively for each of the models tested. That
 671 is why there are multiple circles for components of the base model because the base model was
 672 part of all the models test (that is, one model for each personality trait).

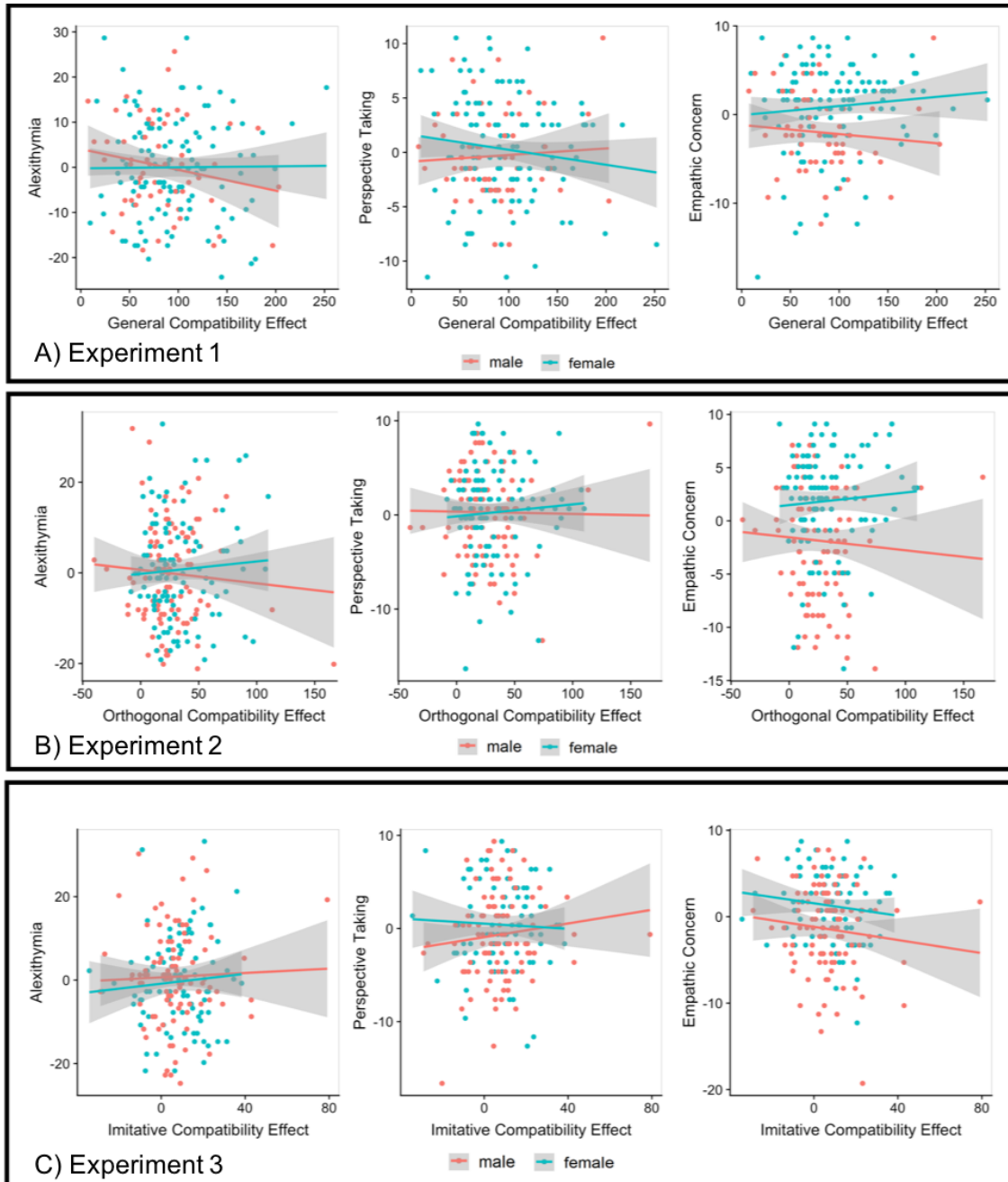
673
 674 To evaluate the sex*trait interaction terms, we computed additional models –
 675 each model consisted of the base model, one trait predictor, and the sex*trait

676 interaction term. For alexithymia, when the sex*trait term was included in the model,
677 the model explained 34.9% of the variance. The sex*alexithymia interaction term
678 marginally predicted the compatibility effect ($B=0.48$, $SEB=0.27$, $t(169)=-1.71$, $p=0.09$,
679 95% CI [-0.07, 1.03]), and explained an additional 1.3 % of the variance ($\Delta R^2= .013$,
680 $F(1,169)=2.93$, $p<0.09$). A decrease in alexithymia marginally predicted an increase in
681 the compatibility effect only for males ($B=-0.80$, $SEB=0.47$, $t(169)=-1.72$, $p=0.09$) and
682 not for females ($B=0.15$, $SEB=0.30$, $t(169)=0.49$, $p=0.62$; see Figure 6A). The effect size
683 attributable to the addition of alexithymia and the sex*trait interaction term was very
684 small (Cohen's $f^2 = 0.02$).

685 For empathy, when the sex*trait term was included in the model, the model explained
686 36.9% of the variance. Sex*empathic concern predicted the compatibility effect above
687 and beyond the base model ($B=1.62$, $SEB=0.74$, $t(167)=2.17$, $p=0.03$, 95% CI [0.15,
688 3.10]) and explained an additional 1.8 % of the variance ($\Delta R^2= .018$, $F(1,167)=4.73$,
689 $p=.031$). Sex*perspective taking ($B=-1.91$, $SEB=0.81$, $t(167)=-2.34$, $p=0.02$, 95% CI [-
690 3.52, -0.30]) predicted the compatibility effect above and beyond the base model and
691 explained an additional 2.1 % of the variance ($\Delta R^2= .021$, $F(1,167)=5.49$, $p=.021$). An
692 increase in empathic concern marginally predicted a decrease in the compatibility effect
693 for males ($B=-2.24$, $SEB=1.23$, $t(167)=-1.82$, $p=.07$) whereas in females, there was a
694 trend for an increase in empathic concern predicting an increase in the compatibility
695 effect ($B=1.01$, $SEB=0.85$, $t(169)=1.19$, $p=.24$). An increase in perspective taking
696 predicted a decrease in the compatibility effect in females ($B=-1.90$, $SEB=0.85$, $t(167)=-$
697 2.24 , $p=.026$). In males, there was a trend for an increase in perspective taking
698 predicting an increase in the compatibility effect ($B=1.92$, $SEB=1.39$, $t(167)=1.38$,
699 $p=0.169$; see Figure 6A). The effect size attributable to the addition of empathy and the
700 sex*trait interaction term was small (Cohen's $f^2 = 0.05$). None of the other sex*trait

701 interaction terms predicted the compatibility effect above and beyond the base model
702 (Cohen's f^2 for all models <0.02 , Supplementary Figure 5, Supplementary Table 5).

703



704 **Figure 6.** Sex by trait interactions for Experiments 1, 2, 3. Sex by trait interactions for alexithymia,
705 perspective taking, and empathy personality traits for Experiments 1 (A), 2 (B), and 3 (C). X axis
706 denotes the imitative compatibility effect in milliseconds, and Y axis denotes mean centred scores on
707 the personality traits.
708

709

710

711 **Discussion**

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

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

733 A second limitation to these initial conclusions is that the flanker task used in
734 the current experiment employed fewer trials than those used in previous studies

735 where a sex difference was found (e.g. Stoet, 2011; Clayson et al., 2011). Therefore, a
736 lack of sex difference might reflect a lack of precision in measuring the effect. This might
737 also explain why we did not find a main effect of compatibility on the accuracy data in
738 the flanker task. Thus, although the current experiment employed a larger sample size
739 than previous studies using the flanker task, we are still cautious to interpret the lack of
740 evidence for the sex difference in the first experiment.

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

748 Further, in the current experiment, we did not find any effect of the type of
749 emotional expression on automatic imitation. These findings add to previous research
750 that shows mixed evidence for a link between the emotional expression of the
751 interacting partner and the tendency to automatically imitate (Grerucci et al., 2013;
752 Crescentini et al., 2011; Rauchbauer et al., 2015). Finally, if we turn to consider the
753 effects of stable personality measures, a clear picture begins to emerge. Even in a more
754 socially meaningful context where emotional expressions are signalled, we further
755 support the claim that imitative control in general (across the entire group of
756 participants), shows a general invariance to stable dimensions of personality like
757 narcissism, agreeableness, extraversion, autistic-like and schizotypal traits (Butler et al.,
758 2015; Cracco et al., 2018). Of course, it is possible that the emotional expressions failed
759 to add to the social context of the task in a meaningful way. The face image signalling

760 the emotional expression was presented along with the hand image but they were two
761 separate images. This ensured that participants also paid attention to the hand image
762 which was the focus of the task. Further, it has been recently suggested that the link
763 between inferring an emotional state and the corresponding facial movements or
764 expressions is less clear than what has been previously suggested (Barrett et al., 2019).
765 It is possible, therefore, that participants ignored the face image, did not think of the
766 hand as connected to the face image, and did not infer the emotional state of the stimuli.
767 Even if this were true, however, we add a further large dataset to the prior work (Butler
768 et al., 2015; Cracco et al., 2018), which all show that personality variables have little
769 relationship to performance on the imitation task in general. We suggest that studies
770 purporting alternative patterns of relationship between imitation and personality
771 measures in general across the population (Chartrand & Bargh, 1999; Obhi et al., 2013;
772 Hogeveen & Obhi, 2013) perform powerful replications to enable a cumulative science
773 to develop (Munafò et al., 2017; Zwaan et al., 2018).

774 Although there were no clear main effects of personality across the entire
775 group, there was some suggestive evidence that the effect of personality on imitation
776 differed by sex. Given prior evidence linking automatic imitation and alexithymia, we
777 expected that an increase in automatic imitation would be predicted by a decrease in
778 alexithymia (Sowden et al., 2016). In the current experiment, this was true only for
779 males, and not for females. We further predicted that an increase in empathic concern
780 and perspective taking would predict an increase in automatic imitation (Chartrand &
781 Bargh, 1999). However, the current findings suggest that a decrease in empathic
782 concern predicts the compatibility effect in males, but not females, and a decrease in
783 perspective taking predicts the compatibility effect in females, but not males. It has been
784 suggested that males score higher on measures of alexithymia as compared to females

785 (Levant et al., 2009), and females score higher on empathic concern and perspective
786 taking as compared to males (Christov-Moore et al., 2014; Van der Graaff et al., 2013).
787 We had no *a priori* hypotheses however as to whether females and males would show a
788 link between personality and imitation in different directions. In addition, these
789 sex*trait interactions were small effects and contributed to only an additional 1.3%
790 (alexithymia) and 3.3 % (empathy) of the model. Thus, before making any firm
791 conclusions, these results require replication in order to confirm that they do not reflect
792 false positives as a result of sampling error.

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

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

805

806

807

808

809

Experiment 2

810 **Introduction**

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

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

827 **Method**

828 *Participants*

829 Two hundred and thirty-eight participants took part in this experiment for
830 monetary compensation (£6) or course credit. All participants provided informed
831 consent, had normal or corrected-to-normal vision, and were right-handed. Approval
832 was obtained from the Research Ethics and Governance Committee of the School of

833 Psychology at Bangor University. One participant was excluded because data on only
834 half the trials was recorded on the flanker task.

835 Participants were excluded if performance was three standard deviations away
836 from the group mean average performance per condition in terms of accuracy or
837 reaction time (N=15 for the imitation task, N=21 for the flanker task). For the imitation
838 task, the final sample included 223 participants (107 males, Mean_{age} = 20., SD_{age} = 4.33;
839 Mean_{age} and SD_{age} are based on 203 participants as some participants did not enter their
840 age in the demographic questionnaire). For the flanker task, the final sample included
841 217 participants (101 males, Mean_{age} = 20.7, SD_{age} = 4.31; Mean_{age} and SD_{age} are based
842 on 198 participants).

843 *Stimuli, tasks, and procedure*

844 *Automatic imitation task.* The automatic imitation task was similar to the one
845 used in Experiment 1, with the following changes: one, no face image was presented
846 during the task (Figure 1). Two, the hand stimuli were presented orthogonal to the
847 response (Figure 1). Three, there were 360 trials in total, which comprised six blocks of
848 60 trials, each of which included 30 compatible and 30 incompatible trials.

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

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

854 *Questionnaires.* Participants also completed two self-report questionnaires which
855 included the Interpersonal Reactivity Index (IRI; Davis, 1980), and the Toronto

856 Alexithymia Scale (TAS-20; Bagby et al., 1994). For more details on the measures used,
857 see the Supplementary Material.

858 **Data Analysis**

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

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

870 **Results**

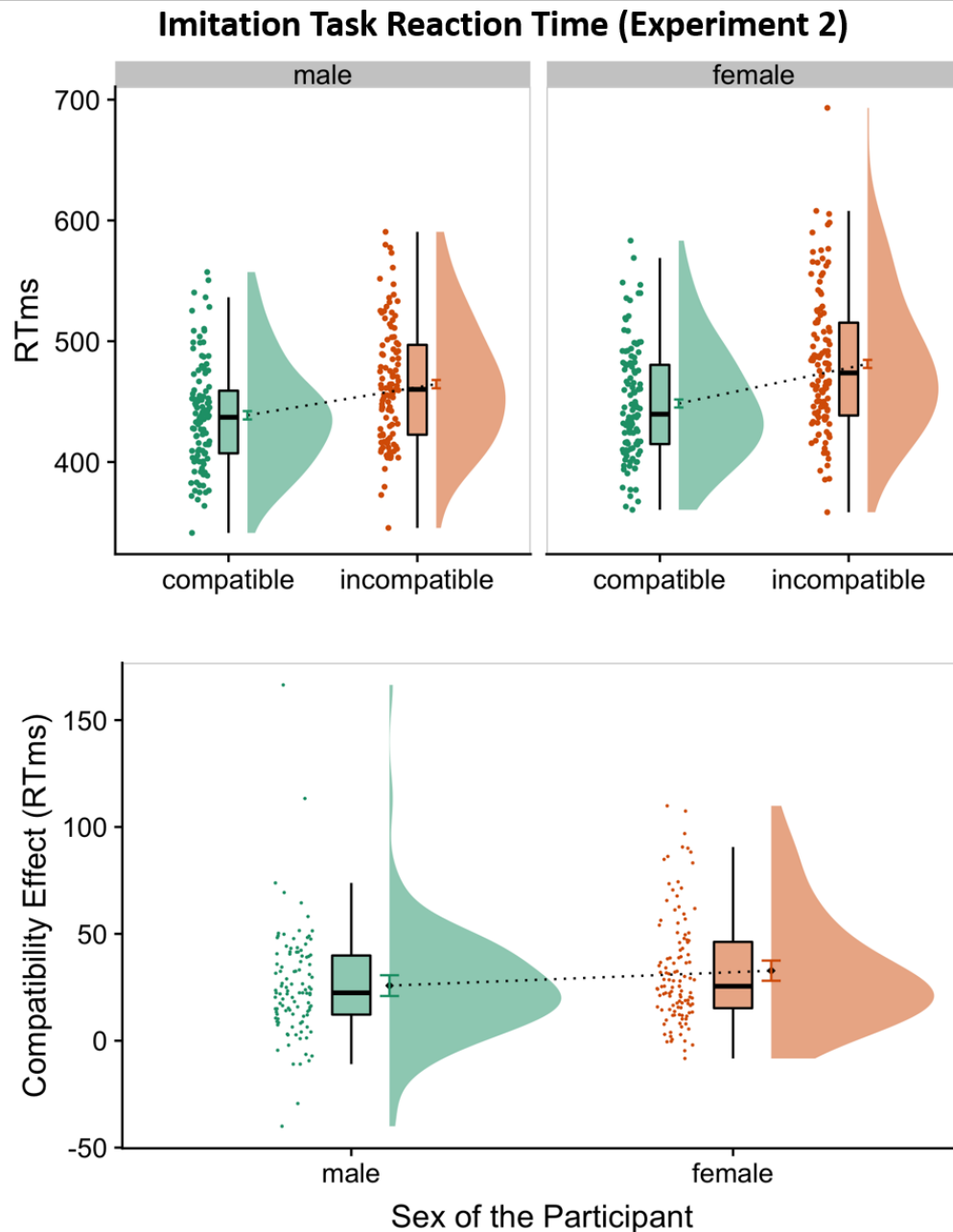
871 *Automatic imitation task*

872 *Accuracy.* Average accuracy for both males and females for both compatible and
873 incompatible trials was over 92% (see Supplementary Figure 7, Supplementary Table
874 6). The 2 (compatibility: compatible, incompatible) x 2 (sex: male, female) mixed
875 ANOVA showed a main effect of compatibility such that participants were more
876 accurate on compatible trials than incompatible trials ($F(1, 221) = 96.22, p < .001, \eta_p^2 =$
877 0.30). The effect size for the main effect of compatibility is conventionally considered to
878 be large. The effect sizes for the main effect of sex ($F(1, 221) = 1.87, p = 0.17, \eta_p^2 = 0.008$)

879 and the Compatibility*Sex interaction ($F(1, 221) = 0.14, p = 0.71, \eta_p^2 < 0.001$) were close
880 to zero with p-values > 0.1 (see Supplementary Table 7).

881 *Reaction time.* Mean reaction times were between 435 and 485 milliseconds for
882 both males and females on both compatible and incompatible trials (see Figure 7,
883 Supplementary Table 6). The 2 (compatibility: compatible, incompatible) x 2 (sex: male,
884 female) mixed ANOVA (Figure 7) showed a main effect of compatibility such that
885 participants were slower to respond on incompatible trials than compatible trials ($F(1,$
886 $221) = 293.18, p < .001, \eta_p^2 = 0.56$). The effect size for the main effect of compatibility is
887 conventionally considered to be large. The main effect of sex had a relatively small
888 effect size and showed that females were generally slower than males ($F(1, 221) = 4.23,$
889 $p = 0.040, \eta_p^2 = 0.02$). There was a Compatibility*Sex interaction and the effect size is
890 conventionally considered to be a small effect ($F(1, 221) = 4.17, p = 0.042, \eta_p^2 = 0.02;$
891 Supplementary Table 7).

892 In order to interrogate our primary hypothesis regarding sex differences in the
893 imitation task, we computed compatibility effects separately for males and females, and
894 then compared them to each other. For both males and females, compatibility effects
895 had a large effect size (Cohen's $d_z > 1.0$) and the lower bound of the 95% confidence
896 interval was at least 0.82. When compatibility effects for males and females were
897 directly compared to each other, we found a mean difference of 6.98 ms in the direction
898 that was predicted i.e. the compatibility effect for females was greater than the
899 compatibility effect for males with the lower bound of the 95% confidence interval
900 above zero (Mean Difference = 6.98 ms, 95% CI[1.34, ∞], Cohen's $d = 0.27$ 95% CI[0.05,
901 ∞]; Figure 7, Table 2B). The effect size was a small-to-medium effect, with the lower
902 bound of the 95% CI at 0.05.



903 **Figure 7.** *Experiment 2 – Imitation Task Reaction Time.* Reaction time is reported in milliseconds
 904 (ms). The upper panel shows mean reaction times for compatible and incompatible conditions
 905 for both males and females. The lower panel shows the compatibility effect for both males and
 906 females. The compatibility effect is calculated by subtracting reaction times on compatible trials
 907 from incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms =
 908 reaction time in milliseconds.
 909

910

911 The absolute size of the difference between the sexes as measured in original
 912 units (i.e., ms) is smaller than Experiment 1, as the orthogonal compatibility effect is
 913 smaller than the general compatibility effect measured in Experiment 1. Indeed, when

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

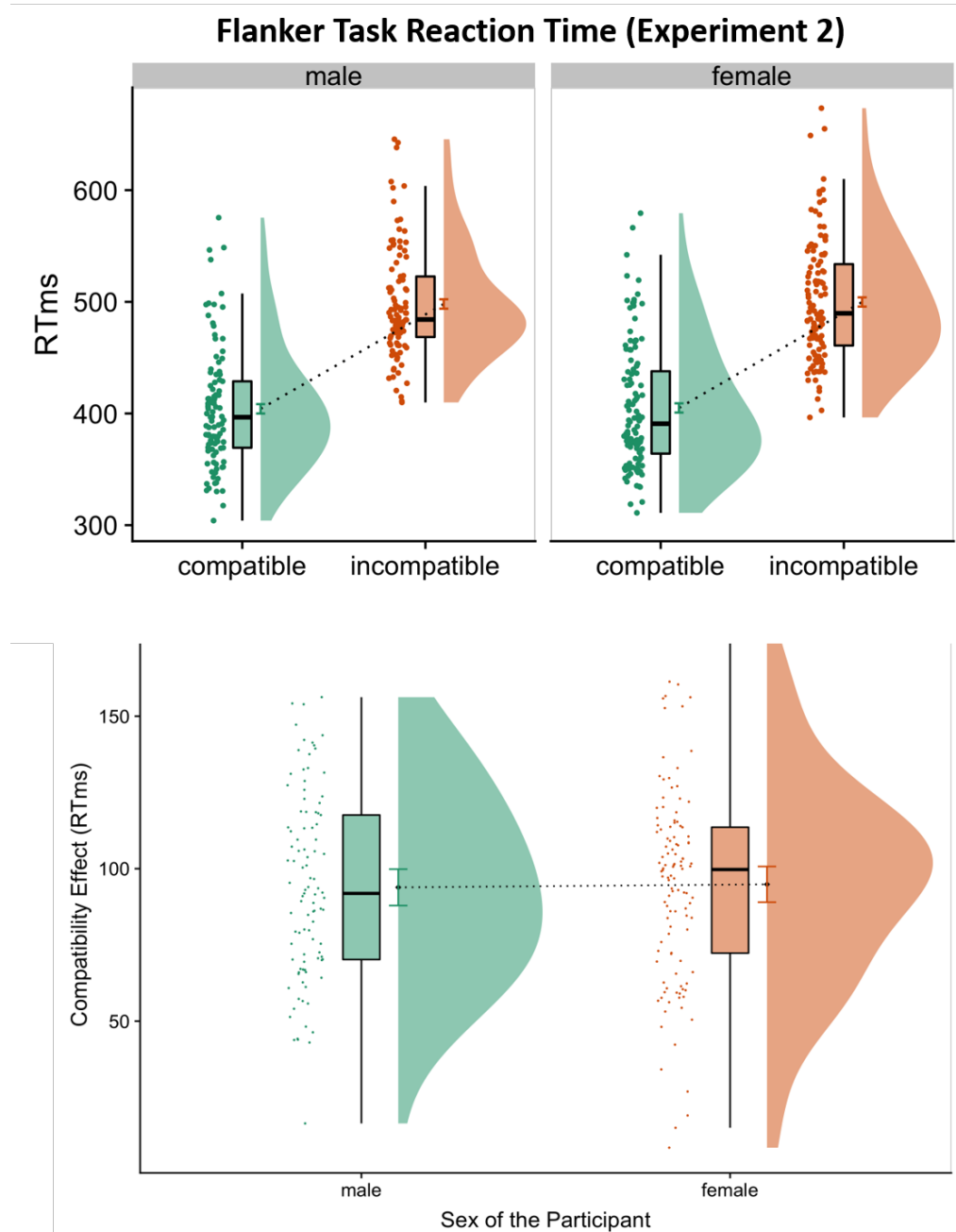
924 *Flanker task*

925 *Accuracy.* Average accuracy was over 88% for both males and females on both
926 compatible and incompatible trials (see Supplementary Figure 8, Supplementary Table
927 6). A 2 (compatibility: compatible, incompatible) x 2 (sex: male, female) mixed ANOVA
928 showed a main effect of compatibility ($F(1, 215) = 151.33, p < .001, \eta_p^2 = 0.41$). The main
929 effect of compatibility showed that participants were more accurate on compatible
930 trials compared to incompatible trials. The main effect of sex showed that females had
931 lower accuracy overall compared to males ($F(1, 215) = 5.78, p = 0.017, \eta_p^2 = 0.03$). The
932 Compatibility*Sex interaction ($F(1, 215) = 3.17, p = 0.076, \eta_p^2 = 0.01$) showed that the
933 difference in accuracy between compatible and incompatible trials was greater for
934 females compared to males. The effect sizes for both the main effect of sex, and the
935 interaction were relatively small (see Supplementary Table 7).

936 *Reaction time.* Mean reaction times for both males and females for both
937 compatible and incompatible conditions was between 400 and 500 milliseconds (see

938 Figure 8, Supplementary Table 6). A 2 (compatibility: compatible, incompatible) x 2
939 (sex: male, female) mixed ANOVA (Figure 8) showed a main effect of compatibility such
940 that participants were slower to respond on incompatible trials than compatible trials
941 ($F(1, 215) = 1986.89, p < .001, \eta_p^2 = 0.90$). The effect size for the main effect of
942 compatibility is conventionally considered to be large. The effect sizes for the main
943 effect of sex ($F(1, 215) = 0.03, p = 0.854, \eta_p^2 < 0.001$), and the Compatibility*Sex
944 interaction ($F(1, 215) = 0.05, p = 0.816, \eta_p^2 < 0.001$) were close to zero with p-values >
945 0.8 (see Supplementary Table 7).

946 To explore sex differences in the flanker task further, compatibility effects were
947 computed separately for males and females, and then compared to each other. For both
948 males and females, compatibility effects had a large effect size (Cohen's $d_z > 2.9$) with
949 the lower bound of the 95% confidence interval at least 2.62. When compatibility effects
950 for males and females were directly compared to each other, females showed a higher
951 compatibility effect than males, but the effect size was very small, with the lower bound
952 of the 95% confidence interval reaching -6.01ms (Mean Difference = 0.98 ms, 95% CI[-
953 6.01, ∞], Cohen's $d = 0.03, 95\%CI[-0.19, \infty]$; Figure 8, Table 2B). The effect size was
954 close to zero with the lower bound of the confidence interval at -0.19 (below zero). A
955 Bayesian independent samples t-test showed that the null was 6 to 7 times more likely
956 than the alternative hypothesis ($BF_{01} = 6.58$). Thus, although both males and females
957 separately showed a compatibility effect, there was a negligible difference between
958 males and females on the flanker compatibility effect.



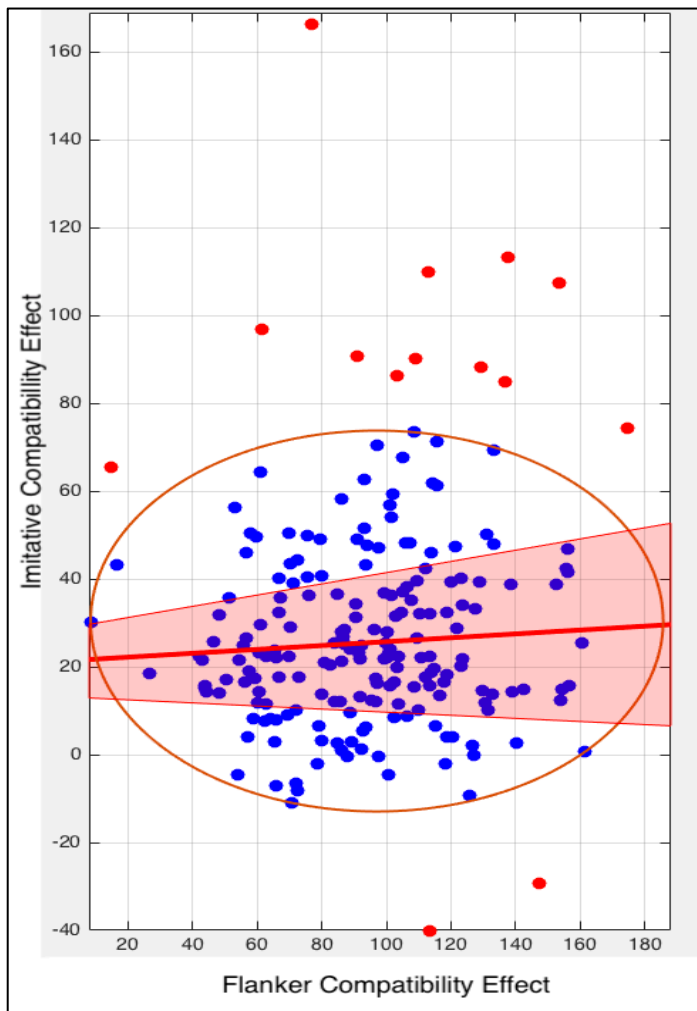
959 **Figure 8.** *Experiment 2 – Flanker Task Reaction Time.* Reaction time is reported in milliseconds
 960 (ms). The upper panel shows mean reaction times for compatible and incompatible conditions
 961 for both males and females. The lower panel shows the compatibility effect for both males and
 962 females. The compatibility effect is calculated by subtracting reaction times on compatible trials
 963 from incompatible trials. Error bars represent 95% confidence intervals. Abbreviations: RTms =
 964 reaction time in milliseconds.
 965

966

967 *Correlational analysis*

968 In order to investigate whether the flanker and imitation compatibility effects
 969 were correlated, a one-tailed skipped correlation was performed. For the correlational

970 analysis, only those participants who performed both the tasks were included in the
971 analysis (N=205). As in Experiment 1, we also performed a more robust correlation
972 analysis. The data was not normally distributed, but was homoscedastic. Thus, we
973 performed a skipped Spearman correlation analysis on 191 participants as 14 bivariate
974 outliers were detected. Results indicated that flanker and imitation compatibility effects
975 showed a weak positive correlation that did not pass our statistical threshold
976 (Spearman $r(191)=0.07$, 95% CI $[-0.07, \infty]$; Figure 9). Our findings thus suggest that
977 flanker and imitative compatibility effects are largely unrelated, and interference on one
978 task did not predict interference on the other.

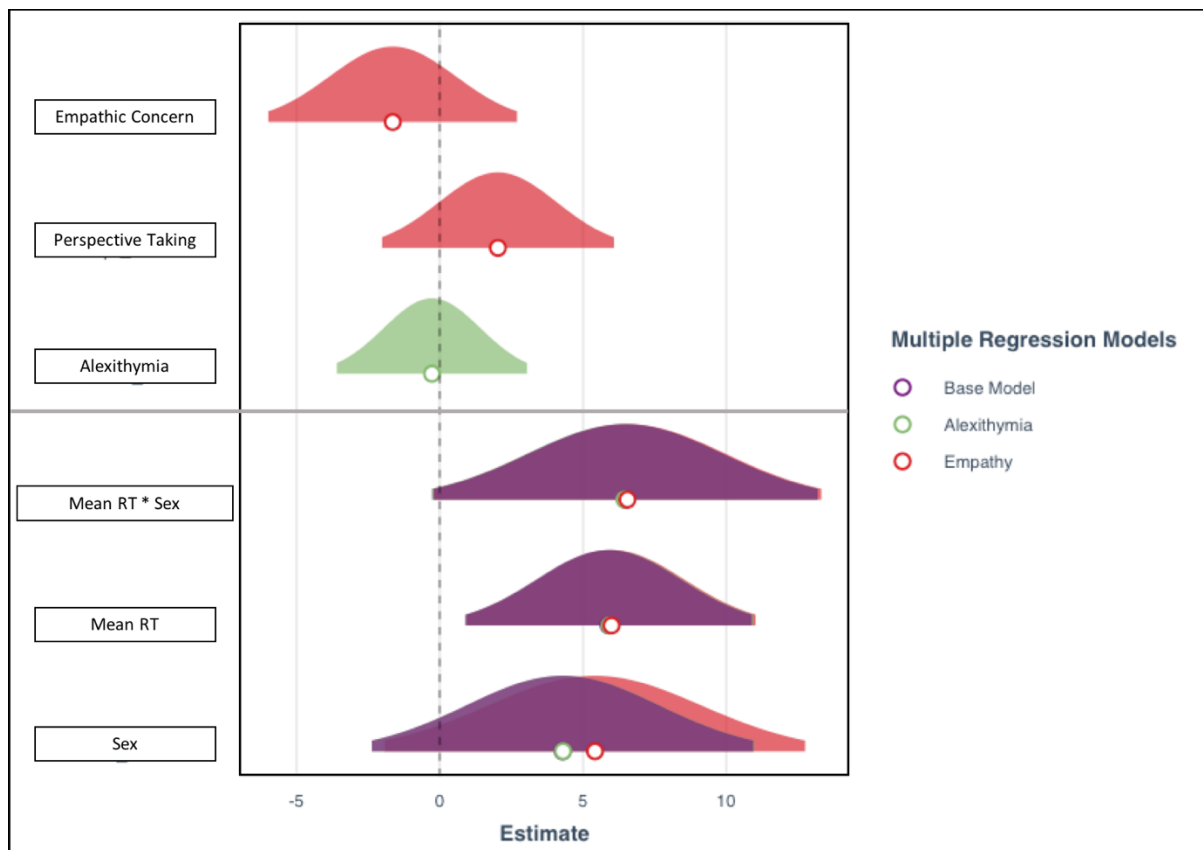


979 **Figure 9.** Experiment 2 - Correlation Analysis. A skipped Spearman correlation shows a small
980 positive correlation between the flanker and imitative compatibility effects that does not pass
981 our statistical thresholding. Abbreviations: RTms (reaction time in milliseconds). Dots in red are
982 the bivariate outliers.
983

984 *Multiple regression analyses*

985 We also investigated the relationship between personality variables (empathy
986 and alexithymia) and the orthogonal compatibility effect as measured on the SRC task.
987 Tests for multicollinearity indicated that a very low level of multicollinearity was
988 present (*VIF* for all predictor variables < 2). The base model (which included mean RT,
989 sex, and the mean RT*Sex interaction) explained 16.2% of the variance in the
990 congruency effect ($F(3,204)=13.13$, $p<.001$, $R^2=.16$, $f^2 = 0.19$) and indicated a medium
991 effect. Mean RT predicted the orthogonal compatibility effect ($B=0.19$, $SEB=0.03$,
992 $t(204)=5.36$, $p<.001$, $[0.12; 0.26]$) with increasing CE as mean RT increased. Sex did not
993 predict the orthogonal compatibility effect ($B=2.14$, $SEB=1.67$, $t(204)=1.27$, $p=.21$, $[-$
994 $1.19; 5.47]$). The mean RT * sex interaction ($B=0.07$, $SEB=0.03$, $t(204)=1.90$, $p=.06$, $[-$
995 $0.002; 0.13]$) suggested that increases in mean RT predicted larger increases in the
996 compatibility effect for females ($B=0.25$, $SEB=0.05$, $t(204)=5.45$, $p<.001$) compared to
997 males ($B=0.12$, $SEB=0.05$, $t(204)=2.33$, $p<.001$).

998 Alexithymia and empathy (empathic concern and perspective taking) did not
999 predict the orthogonal compatibility effect above and beyond the base model (all p 's >
1000 0.03, all CIs overlapping with zero; see Figure 10). Effect sizes attributable to the
1001 addition of the personality variables (beyond the base model) indicated extremely small
1002 effects (Cohen's $f^2 = <.001$ for alexithymia and Cohen's $f^2 = .005$ for empathy). The
1003 multiple regression models are summarized in Supplementary Table 8. Zero-order
1004 correlations are also consistent with the findings from the multiple regression analyses
1005 (see Supplementary Table 9, Supplementary Figure 9).



1006
 1007 **Figure 10.** *Experiment 2 – Multiple Regression Analyses.* Values of standardised coefficients are
 1008 plotted for each predictor variable (personality trait) along with their corresponding
 1009 uncertainties (95% confidence interval width for a normal distribution for each estimate).
 1010 Coefficients are standardised by dividing by two standard deviation units according to Gelman
 1011 (2008). The base model consists in the bottom three predictor variables (depicted in violet) –
 1012 mean RT, Sex, and meanRT*Sex. Abbreviations: RT = Reaction Time. *N.B.* The circles on the
 1013 purple distributions represent the standardised co-efficients for the main effect of meanRT,
 1014 main effect of sex, and meanRT*sex interaction respectively for each of the models tested. That
 1015 is why there are multiple circles for components of the base model because the base model was
 1016 part of all the models test (that is, one model for each personality trait).

1017

1018 To evaluate the sex*trait interaction terms, we computed additional models –
 1019 each model consisted of the base model, one trait predictor, and the sex*trait
 1020 interaction term. None of the sex*trait interaction terms predicted the orthogonal
 1021 compatibility effect above and beyond the base model (all $p > 0.3$, all CIs overlapping
 1022 with zero; Figure 6B, Supplementary Figure 10). Effect sizes attributable to the addition
 1023 of the sex*trait interaction terms (beyond the base model) indicated extremely small
 1024 effects (Cohen's $f^2 = 0.01$ for both alexithymia and empathy). The multiple regression
 1025 models are summarized in Supplementary Table 10.

1026

1027 **Discussion**

1028 As in Experiment 1, results indicated a clear sex difference in the orthogonal
1029 compatibility effect such that females showed a higher orthogonal compatibility effect
1030 compared to males on the automatic imitation task. The sex difference persisted on the
1031 imitation task in spite of presenting stimuli orthogonal to the response. However, this
1032 sex difference on the RT compatibility effect was not found on the flanker task even
1033 after increasing the number of trials. Further, the correlational analysis suggested that
1034 flanker and orthogonal compatibility effects were only marginally correlated with each
1035 other and explained only 0.8% of the variance. Thus, greater interference on one task is
1036 able to predict only a very small amount of interference on the other task.

1037 Thus, across Experiments 1 and 2, we show a lack of consistent evidence for a
1038 sex difference in the flanker task. However, the interpretation of the sex difference on
1039 the imitation task still has two potential limitations. One, while the presentation of
1040 orthogonal stimuli reduces spatial compatibility effects on the left-right axis, they do not
1041 rule out the possibility of orthogonal spatial compatibility effects i.e. the propensity of
1042 participants to show an advantage for an up-right and down-left pairing (Weeks &
1043 Proctor, 1999; Cho & Proctor, 2003; Weeks, Proctor, & Beyak, 1994). For instance, when
1044 stimuli were presented orthogonal to the response hand (see Figure 1), the index finger
1045 was always below the middle finger, and the participant's index finger was to the left
1046 side of space. Thus, a preference for responding to "up" stimuli with a right response
1047 and "down" stimuli with a left response may be observed along with imitative effects on
1048 the automatic imitation task used in the current experiment. Thus, the sex difference

1049 may reflect a difference on orthogonal spatial effects as opposed to purely imitative
1050 effects.

1051 Two, the stimuli used in both Experiments 1 and 2 were those of a female
1052 model. We did not manipulate the sex of the stimulus, and therefore, the sex difference
1053 can either reflect a genuine difference between males and females, or an in-group bias.
1054 A difference between male and female participants (irrespective of the sex of the
1055 stimulus) would reflect distinct (or partially distinct) cognitive mechanisms underlying
1056 imitative or spatial control as a function of sex. On the contrary, an in-group bias or
1057 own-sex bias would suggest that sex differences as evidenced previously on the
1058 automatic imitation task (Butler et al., 2015; Genschow et al., 2017) do not actually
1059 reflect a sex difference - females show a higher compatibility effect because they favour
1060 members of the in-group i.e. of their own sex compared to members of the out-group i.e.
1061 of the opposite sex (Rudman & Goodwin, 2004; Brown, 1995; Rauchbauer et al., 2015;
1062 Gleibs et al., 2016).

1063 Finally, in terms of personality measures, empathy and alexithymia (and
1064 sex*trait interactions) did not modulate the orthogonal compatibility effect. Although
1065 we found suggestive evidence in Experiment 1 for a small link between personality
1066 traits (alexithymia and empathy) and imitation that differed between the sexes, the
1067 current experiment did not replicate these findings. Therefore, overall, these results
1068 provide limited support for a link between personality traits and automatic imitation,
1069 and confirm and replicate findings from previous large sample studies (Butler et al.,
1070 2015; Cracco et al., 2018) that suggest automatic imitation is largely invariant to stable
1071 traits of personality.

1072

1073 **Experiment 3**

1074 **Introduction**

1075 Experiment 3 addressed two remaining issues. First, we measured the imitative
1076 compatibility effect independently from the spatial compatibility effect, in order to
1077 estimate whether the sex difference reflects a spatial or more specialised (social)
1078 aspects of cognitive control. Second, we assessed the extent to which the sex difference
1079 reflects a basic difference between males and females and/or an in-group bias based on
1080 sex.

1081 To separate imitative and spatial components of the task, we used a modified
1082 version of the SRC task of automatic imitation that allowed us to manipulate imitative
1083 and spatial effects separately (Catmur & Heyes, 2011; Boyer et al., 2012; Bertenthal et
1084 al., 2006). A sex difference on spatial compatibility alone would indicate that the sex
1085 difference observed in Experiments 1 and 2 can be explained by differences associated
1086 with processing spatial information. Alternatively, a sex difference on imitative
1087 compatibility alone, would suggest that greater compatibility effects for females reflects
1088 a difference in the control of automatic imitation specifically.

1089 To compare a sex difference account with an in-group bias account of our
1090 findings so far, we manipulated the sex of the stimuli used in the SRC task and again
1091 tested male and female participants. A greater compatibility effect for females for
1092 female stimuli compared to male stimuli would indicate that an own-sex bias
1093 contributes to the sex difference observed on the automatic imitation task.
1094 Alternatively, a sex difference on the task and relative invariance to the sex of the
1095 stimuli would suggest that there is a basic control mechanism that differs between

1096 males and females that seems resistant to possible contextual factors, such as group
1097 biases.

1098 In order to investigate whether personality variables influence automatic
1099 imitation, in Experiment 3, we included all personality variables included in Experiment
1100 1 (alexithymia, empathy, autistic-like and schizotypal traits, narcissism, extraversion,
1101 and agreeableness). In Experiment 1 and 2, the compatibility effect measured on the
1102 imitation task was a composite of spatial and imitative effects. Therefore, the invariance
1103 of the compatibility effect may be related to spatial effects as opposed to imitative
1104 effects. Therefore, we included all the personality measures in order to investigate
1105 whether imitative compatibility when measured independently of spatial effects is also
1106 invariant to stable personality traits.

1107 **Method**

1108 *Participants*

1109 Two hundred and one participants took part in this experiment for monetary
1110 compensation (£6) or course credit. All participants provided informed consent, had
1111 normal or corrected-to-normal vision, and were right-handed. Approval was obtained
1112 from the Research Ethics and Governance Committee of the School of Psychology at
1113 Bangor University. Participants were excluded if performance was 3 standard
1114 deviations away from the group mean average performance per condition in terms of
1115 accuracy or reaction time on the imitation task ($N = 12$). The final sample included 189
1116 participants (97 males, $Mean_{age} = 21.4$, $SD_{age} = 4.08$, age range = 18 to 42) ($Mean_{age}$ and
1117 SD_{age} are based on 182 participants as 7 participants did not enter their age in the
1118 demographics questionnaire).

1119 *Stimuli, tasks, and procedure*

1120 *Automatic imitation task.* The automatic imitation task was similar to the one
1121 used in Experiment 2, with the following changes: one, stimuli were not presented
1122 orthogonally to the response. Two, we calculated an imitative compatibility effect
1123 independent of the spatial compatibility effect (Catmur & Heyes, 2011). For this, both
1124 left- and right-hand images were used as stimuli, but participants always responded
1125 with their right hand. This resulted in eight trial types and four conditions of interest
1126 (Figure 1):

1127 1. imitatively and spatially compatible (for example, when participants are
1128 cued to lift their index finger, and watch an index finger lift of the left hand, the
1129 observed finger movement is both spatially and imitatively compatible to the executed
1130 movement),

1131 2. imitatively and spatially incompatible (for example, when participants are
1132 cued to lift their index finger, and watch a middle finger lift of the left hand, the
1133 observed finger movement is both spatially and imitatively incompatible to the
1134 executed movement),

1135 3. imitatively compatible and spatially incompatible (for example, when
1136 participants are cued to lift their middle finger, and watch a middle finger lift of the
1137 right hand, the observed finger movement is imitatively compatible, but spatially
1138 incompatible to the executed movement),

1139 4. imitatively incompatible and spatially compatible (for example, when
1140 participants are cued to lift their middle finger, and watch an index finger lift of the right
1141 hand, the observed finger movement is imitatively incompatible, but spatially
1142 compatible to the executed movement).

1143 Thus, participants performed the same (imitatively compatible) or different
1144 (imitatively incompatible) movement on the same (spatially compatible) or different
1145 (spatially incompatible) side of space.

1146 A third change in comparison to Experiment 2, is that in order to investigate
1147 whether the sex difference was due to an own-sex bias, the hand stimuli presented
1148 included 4 female and 4 male hands. The hand stimuli were chosen based on a pilot
1149 study. In the pilot study (see Supplementary Material), eighteen hand stimuli were
1150 rated by 51 participants on a scale of 1 to 9 with one being most masculine, 5 being
1151 neutral, and 9 being most feminine. Four hand stimuli rated as most masculine, and four
1152 hand stimuli rated as most feminine were chosen for the current experiment. There
1153 were 360 total trials, with 90 trials per condition. Timing information and pseudo-
1154 randomisation was the same as in Experiment 1 and 2.

1155 *Questionnaires.* Participants also completed a range of self-report questionnaires
1156 which included the Mini International Personality Item Pool (mini IPIP; Donnellan et al.,
1157 2005; the Short Autism Spectrum Quotient (AQ-10 Adult; Baron-Cohen et al., 2001), the
1158 Brief Schizotypal Personality Questionnaire (SPQ-B; Raine & Benishay, 2005), the
1159 Narcissistic Personality Inventory (NPI-16; Ames et al., 2006), the Hypersensitivity
1160 Narcissism Scale (HSNS; Hendin & Cheek, 1997), the Interpersonal Reactivity Index (IRI;
1161 Davis, 1980), and the Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994). For more
1162 details on the measures used, see the Supplementary Material.

1163 In order to confirm that participants perceived male and female stimuli
1164 differently, participants also rated the hand stimuli used in the experiment after they
1165 completed the task. Participants were asked to rate the stimuli on a scale of 1 to 9, with
1166 1 being extremely masculine and 9 being extremely feminine.

1167 *Data Analysis*

1168 Accuracy and RT on the imitation task were recorded in the same way as
1169 Experiment 1 and 2 and only correct trials were used to calculate RT. Trials on which
1170 participants responded incorrectly, i.e. lifted the wrong finger, responded after 2000
1171 ms, or before target onset (7.41 %) were all excluded from the analysis.

1172 A key aim of our study was to investigate whether the sex difference and/or in-
1173 group bias exists in imitative and/or spatial compatibility effects (and not whether/how
1174 such differences differ between the two types of effect). For this purpose, therefore, we
1175 performed analyses separately on the spatial and imitative compatibility effects. For
1176 each compatibility effect separately, we performed a 2 (compatibility: incompatible,
1177 compatible) x 2 (stimulus sex: male hand, female hand) x 2 (participant sex: male,
1178 female) mixed ANOVA on the RT and accuracy data. Based on prior research (Catmur &
1179 Heyes, 2011; Gowen et al., 2016; Marsh et al., 2016; Darda et al., 2018), we expected a
1180 main effect of spatial and imitative compatibility such that RT would be higher, and
1181 accuracy would be lower on spatially incompatible trials compared to spatially
1182 compatible trials, and on imitatively incompatible trials compared to imitatively
1183 compatible trials.

1184 In addition, a Sex*Compatibility interaction for spatial compatibility (such that
1185 females show a higher spatial compatibility effect than males) would be expected if the
1186 sex difference observed in Experiments 1 and 2 was largely driven by the spatial
1187 component of the task. In contrast, a Sex*Compatibility interaction for imitative
1188 compatibility (such that females show a higher imitative compatibility effect than
1189 males) would suggest that the sex difference is largely a reflection of the imitative
1190 component of the task.

1191 Alternatively, if the sex difference in the spatial or imitative compatibility effect
1192 is because of an own-sex bias, we would expect a three-way interaction
1193 (Sex*Compatibility*Stimulus Sex) such that females would be more interfered by a
1194 female stimulus, and males would be more interfered by a male stimulus i.e. females
1195 would show a higher compatibility effect than males for female stimuli compared to
1196 male stimuli.

1197 As in Experiment 1, the interaction effect was central to testing our primary
1198 hypotheses, and thus, we calculated compatibility effects for male and female hand
1199 stimuli separately and independently for both male and female participants. To do so,
1200 we computed the mean difference and 95% confidence intervals between compatible
1201 and incompatible conditions across the levels of stimulus sex and participant sex.
1202 Spatial compatibility was calculated as RT on spatially incompatible trials minus RT on
1203 spatially compatible trials. Imitative compatibility was calculated as RT on imitatively
1204 incompatible trials minus imitatively compatible trials. In order to directly estimate the
1205 size of the difference in spatial and imitative compatibility effects between males and
1206 females, we then again computed the mean differences between the sexes and 95%
1207 confidence intervals. We used one-tailed 95% confidence intervals as we had a
1208 directional hypothesis that females would have a higher spatial or imitative
1209 compatibility effect than males.

1210 For the secondary analyses, multiple regression analyses were performed in the
1211 same way as Experiments 1 and 2 in order to investigate whether personality variables
1212 and sub-clinical traits modulate automatic imitation when measured independent of
1213 spatial effects.

1214

1215 **Results**

1216 *Spatial Compatibility*

1217 *Accuracy.* Average accuracy was over 92% for both males and females for all
1218 conditions of compatibility and stimulus sex (see Supplementary Figure 11,
1219 Supplementary Table 11). A 2 (compatibility: incompatible, compatible) x 2 (stimulus
1220 sex: male hand, female hand) x 2 (participant sex: male, female) mixed ANOVA showed a
1221 main effect of compatibility such that participants were more accurate on compatible
1222 trials than incompatible trials ($F(1,187) = 563.35, p < .001, \eta_p^2 = 0.75$; Supplementary
1223 Figure 11). The effect size of the main effect of compatibility was large. The main effect
1224 of stimulus sex suggested that participants were more accurate when observing male
1225 hand stimuli as compared to female hand stimuli ($F(1,187) = 335.47, p < .001, \eta_p^2 = 0.64$).
1226 The Compatibility*Stimulus Sex interaction suggested that the difference in accuracy
1227 between incompatible and compatible trials was overall bigger for female stimuli
1228 compared to male stimuli ($F(1,187) = 202.31, p < .001, \eta_p^2 = 0.52$), Supplementary Figure
1229 11). All other main effects and interactions' effect sizes were relatively small with high
1230 p-values (see Supplementary Table 12).

1231 *Reaction time.* Mean reaction times for both males and females on all conditions
1232 of compatibility and stimulus sex were between 415 to 475 milliseconds (see Figure 11,
1233 Supplementary Table 11). A 2 (compatibility: incompatible, compatible) x 2 (stimulus
1234 sex: male hand, female hand) x 2 (participant sex: male, female) mixed ANOVA (Figure
1235 11) showed a main effect of compatibility such that participants were slower to respond
1236 on spatially incompatible trials compared to spatially compatible ($F(1,187) = 459.71,$
1237 $p < .001, \eta_p^2 = 0.71$). The effect size of the main effect of compatibility was large. The main
1238 effect of stimulus sex had a medium effect size, and suggested that overall participants

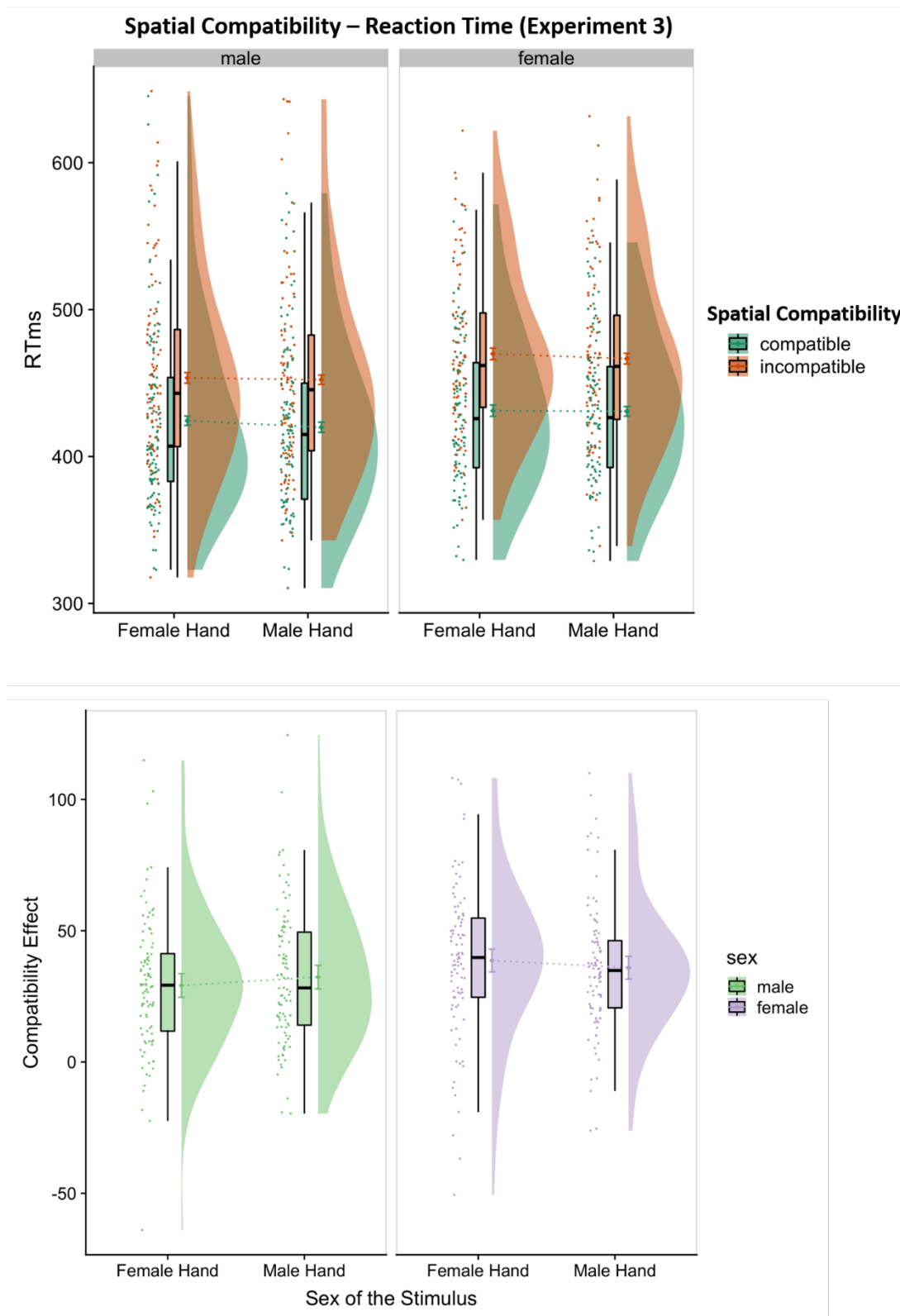
1239 responded slower to female hand stimuli than male hand stimuli ($F(1,187) = 5.63, p =$
1240 $0.019, \eta_p^2 = 0.03$).

1241 The Sex*Compatibility interaction was a small-to-medium effect with a p-value
1242 of 0.04 ($F(1,187) = 4.24, p = 0.041, \eta_p^2 = 0.02$). To interrogate the sex difference in spatial
1243 compatibility, we computed the difference in compatibility effects between males and
1244 females, collapsed across all conditions of stimulus sex. Females showed a higher
1245 compatibility effect than males by 6.53 ms, and the lower bound of the 95% CI was over
1246 zero at 2.02ms (Mean difference = 6.53ms, 95% CI [2.02, ∞], Cohen's $d = 0.24$, 95% CI
1247 [0.07, ∞]; Figure 13A, Table 2C). The effect size was a small-to-medium effect, and the
1248 lower bound of the 95% CI was above zero at 0.07.

1249 The three-way (Compatibility*Stimulus Sex*Sex) interaction ($F(1,187) = 1.77,$
1250 $p=0.185, \eta_p^2 = 0.01$) showed a trend such that females had a higher compatibility effect
1251 for female hand stimuli compared to male hand stimuli, and males had a higher
1252 compatibility effect for male hand stimuli compared to female hand stimuli, although
1253 the effect size was close to zero (Figure 11). All other effect sizes for main effects or
1254 interactions were close to zero (see Supplementary Table 12).

1255 In order to investigate whether an in-group bias explains the sex difference in
1256 spatial compatibility, we computed compatibility effects on all levels of participant sex
1257 and stimulus sex. For both males and females, spatial compatibility effects were present
1258 when observing both male (Cohen's $d_z > 1.25$) as well as female stimuli (Cohen's $d_z >$
1259 1.10). There was a trend for females showing a higher compatibility effect for female
1260 stimuli compared to male stimuli (Mean difference = 2.74 ms, 95% CI [-2.40, ∞],
1261 Cohen's $d_z = 0.09$, 95% CI[-0.08, ∞]), and for males showing a higher compatibility
1262 effect for male stimuli compared to female stimuli (Mean difference = 3.19, 95% CI [-

1263 2.12, ∞], Cohen's $d_z = 0.10$; 95% CI[-0.07, ∞]), but these were relatively small effect
1264 sizes (see Figure 11).



1265

1266 **Figure 11.** *Experiment 3 – Spatial Compatibility Reaction Time.* Reaction time is reported in
1267 milliseconds (ms). The upper panel shows mean reaction times for compatible and

1268 incompatible conditions for both males and females, when responding to both male and female
1269 hand stimuli. The lower panel shows the compatibility effect for both males and females when
1270 responding to both male and female hand stimuli. The compatibility effect is calculated by
1271 subtracting reaction times on compatible trials from incompatible trials. Error bars represent
1272 95% confidence intervals. Abbreviations: RTms = reaction time in milliseconds.

1273

1274 *Imitative Compatibility*

1275 *Accuracy.* Average accuracy for both males and females for all conditions of
1276 stimulus sex and compatibility was above 87% (see Supplementary Figure 12,
1277 Supplementary Table 11). A 2 (compatibility: incompatible, compatible) x 2 (stimulus
1278 sex: male hand, female hand) x 2 (participant sex: male, female) mixed ANOVA
1279 (Supplementary Figure 12) showed a main effect of compatibility ($F(1,187) = 205.65$,
1280 $p < .001$, $\eta_p^2 = 0.52$) such that participants were more accurate on compatible trials than
1281 incompatible trials. The effect size of the main effect of compatibility was large. The
1282 main effect of stimulus sex ($F(1,187) = 335.47$, $p < .001$, $\eta_p^2 = 0.64$) suggested that
1283 participants were more accurate when observing male hand stimuli as compared to
1284 female hand stimuli. The Compatibility*Stimulus Sex interaction ($F(1,187) = 162.98$,
1285 $p < .001$, $\eta_p^2 = 0.46$) suggested that the difference in accuracy between compatible and
1286 incompatible trials was bigger for female stimuli compared to male stimuli (see
1287 Supplementary Figure 12). All other main effects and interactions were relatively small
1288 or close to zero with high p-values (see Supplementary Table 12).

1289 *Reaction time.* Mean reaction times were between 430 and 460 milliseconds for
1290 both males and females on all conditions of compatibility and stimulus sex (see Figure
1291 12, Supplementary Table 11). A 2 (compatibility: incompatible, compatible) x 2
1292 (stimulus sex: male hand, female hand) x 2 (participant sex: male, female) mixed
1293 ANOVA showed a main effect of compatibility ($F(1,187) = 54.96$, $p < .001$, $\eta_p^2 = 0.23$) such
1294 that participants were slower to respond on imitatively incompatible trials compared to

1295 imitatively compatible. The effect size of the main effect of compatibility was large. The
1296 main effect of stimulus sex was a small-to-medium effect with a p-value of 0.02 and
1297 suggested that participants responded slower to female hand stimuli than male hand
1298 stimuli ($F(1,187) = 5.70, p = 0.018, \eta_p^2 = 0.03$).

1299 The effect size for the Sex*Compatibility interaction was close to zero with a p-
1300 value of 0.52 ($F(1,187) = 0.41, p = 0.52, \eta_p^2 = 0.002$). Given the importance to our
1301 primary research question regarding sex differences in the compatibility effect, we
1302 interrogated the RT data further by computing the difference in compatibility effects
1303 between males and females, collapsed across all conditions of stimulus sex. Although
1304 females showed a marginally higher compatibility effect than males by 1.33 ms, the
1305 lower bound of the 95% CI was below zero at -2.15 ms. The effect size was small with
1306 the lower bound of the 95% CI below zero at -0.10 (Mean difference = 1.33, 95% CI [-
1307 2.15, ∞], Cohen's $d = 0.06, 95\% \text{ CI}[-0.10, \infty]$; see Figure 13B, Table 2C).

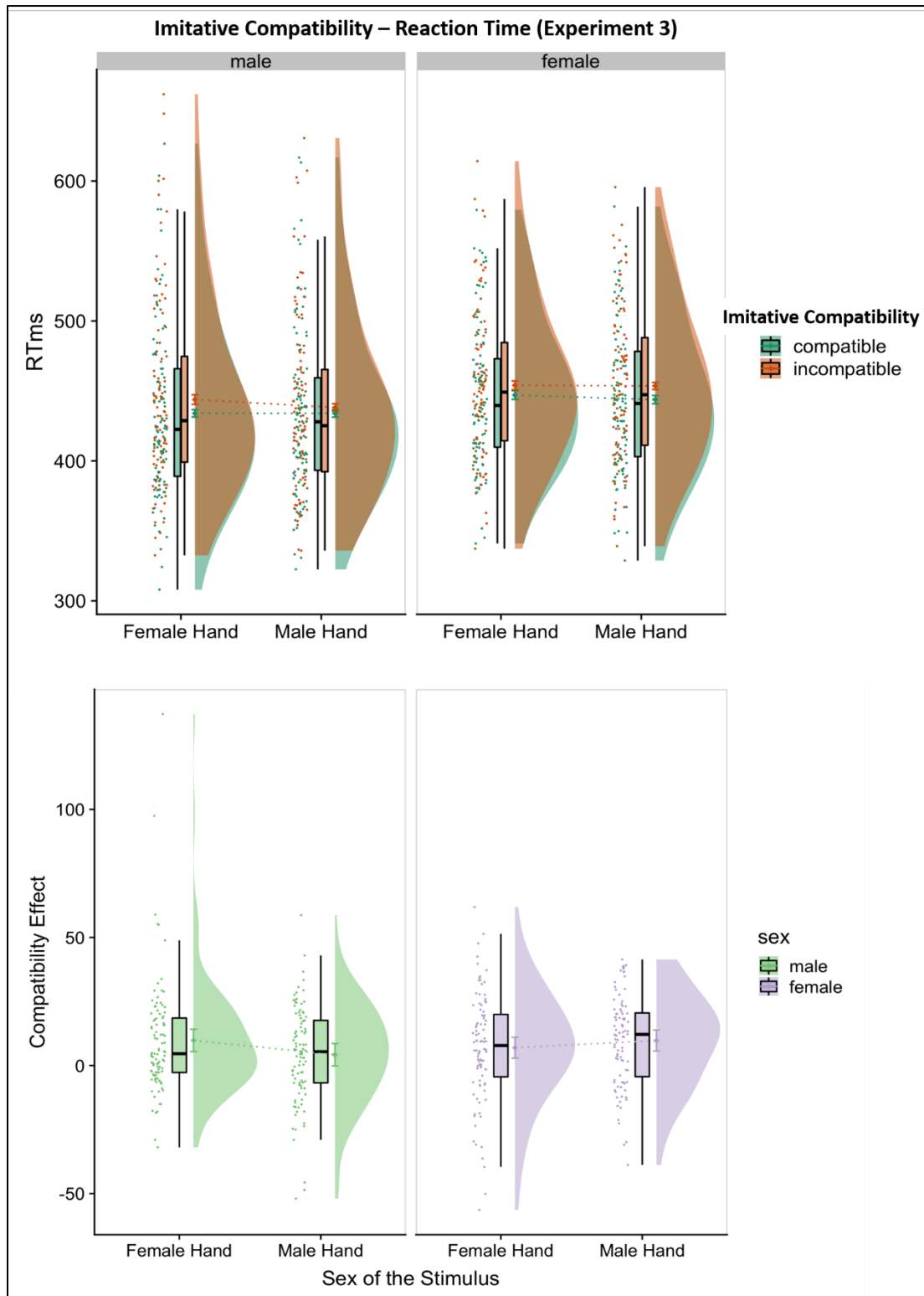
1308 The three-way (Compatibility*Stimulus Sex*Participant Sex) interaction
1309 ($F(1,187) = 3.86, p=0.051, \eta_p^2 = 0.02$) was a relatively small effect. All other main effects
1310 or interactions had effect sizes close to zero with high p-values (see Figure 12,
1311 Supplementary Table 12).

1312 In order to investigate the three-way interaction and explore whether the sex
1313 difference can be explained by an in-group bias, we computed compatibility effects on
1314 all levels of participant sex and stimulus sex. For both males and females, imitative
1315 compatibility effects were present when observing both male (Cohen's $d_z > 0.2$) as well
1316 as female stimuli (Cohen's $d_z > 0.3$). However, there was not even a trend in the
1317 direction we predicted i.e. females did not show a higher compatibility effect for female
1318 stimuli compared to male stimuli (Cohen's $d_z = -0.10$), and for males showing a higher

1319 compatibility effect for male stimuli compared to female stimuli (Cohen's $d_z = -0.18$). On
1320 the contrary, the direction of the interaction was contrary to our hypothesis i.e. females
1321 showed a higher compatibility effect for male stimuli compared to female stimuli, and
1322 males showed a higher compatibility effect for female stimuli compared to male stimuli,
1323 but these effects were small (see Figure 12). As such, not only are these effects relatively
1324 small, they are also inconsistent with the sex difference being a result of an ingroup bias
1325 based on the sex of the interaction partner.

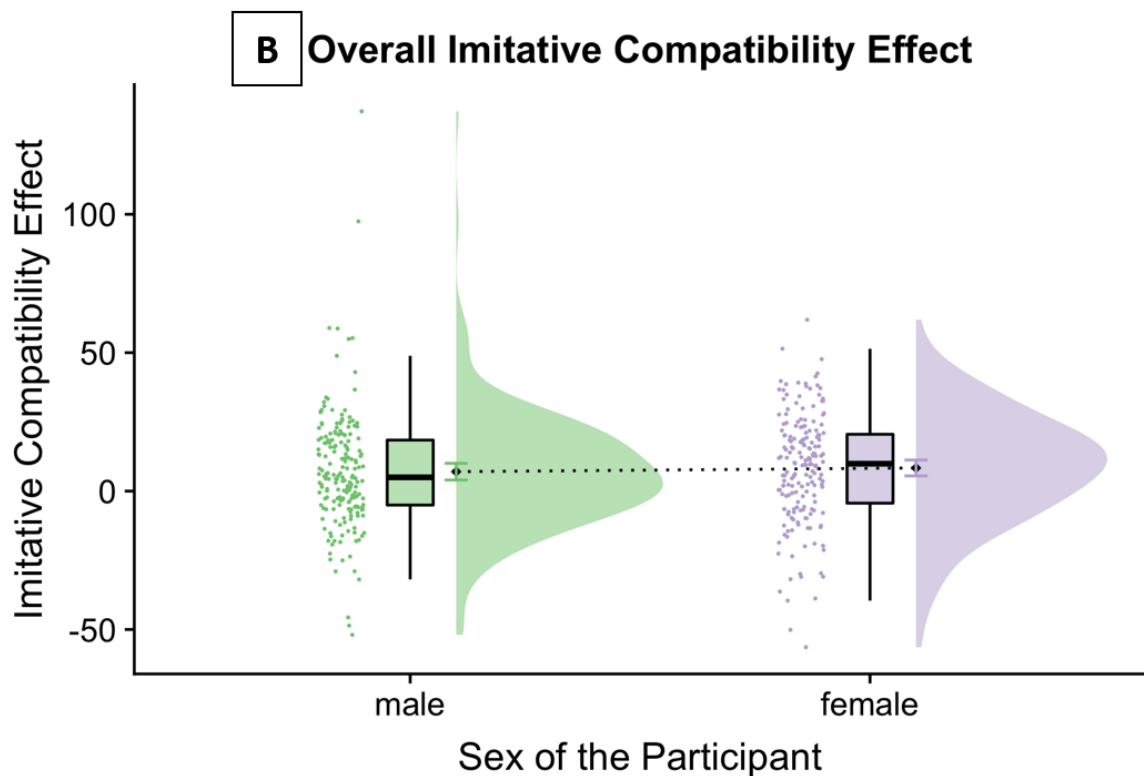
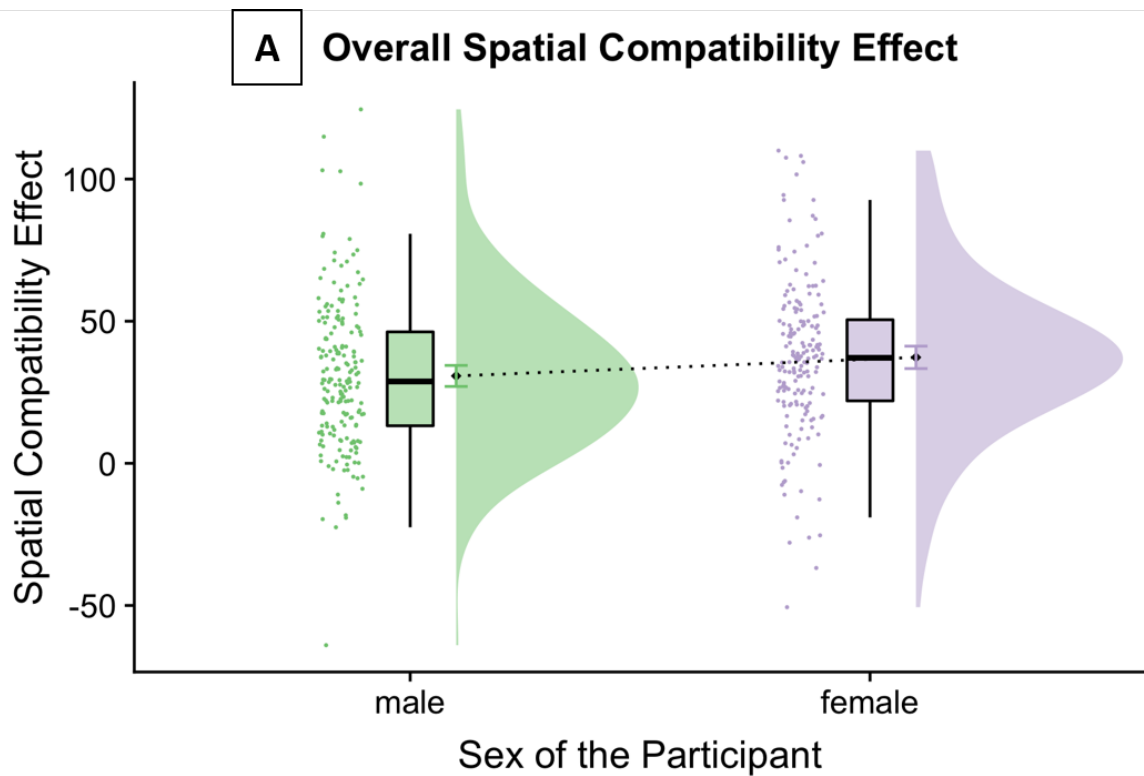
1326 In sum, our results indicated a sex difference in spatial compatibility, but not
1327 imitative compatibility. An in-group bias/own-sex bias did not explain the sex
1328 difference found in the spatial compatibility effect.

1329



1330

1331 **Figure 12.** *Experiment 3 – Imitative Compatibility Reaction Time.* Reaction time is reported in
 1332 milliseconds (ms). The upper panel shows mean reaction times for compatible and
 1333 incompatible conditions for both males and females when responding to both male and female
 1334 hand stimuli. The lower panel shows the compatibility effect for both males and females when
 1335 responding to both male and female hand stimuli. The compatibility effect is calculated by
 1336 subtracting reaction times on compatible trials from incompatible trials. Error bars represent
 1337 95% confidence intervals. Abbreviations: RTms = reaction time in milliseconds.



1338
 1339 **Figure 13.** *Experiment 3 – Overall Compatibility Effects.* The upper panel (A) shows the spatial
 1340 compatibility effect collapsed across sex of the stimulus for both males and females. The lower
 1341 panel (B) shows the imitative compatibility effect collapsed across sex of the stimulus for both
 1342 males and females. The compatibility effect is calculated by subtracting reaction times on
 1343 compatible trials from incompatible trials and is measured in milliseconds. Error bars represent
 1344 95% confidence intervals.

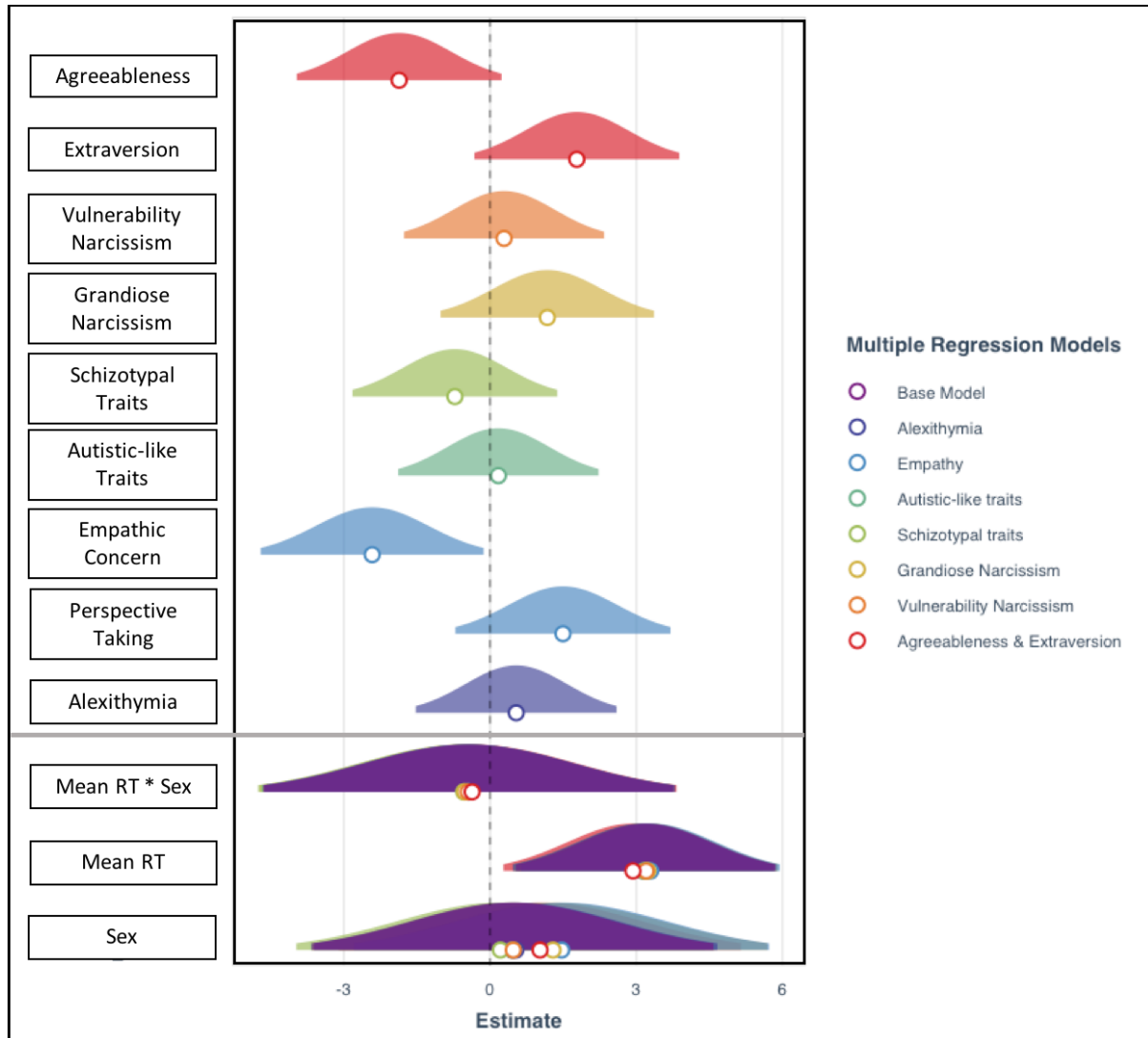
1345

1346 *Multiple regression analyses*

1347 We also investigated the relationship between stable personality measures and
1348 the imitative compatibility effect as measured on the SRC task independent of spatial
1349 effects. Tests for multicollinearity indicated that a very low level of multicollinearity
1350 was present (*VIF* for all predictor variables < 2). The base model (including sex, mean
1351 RT and the sex*mean RT interaction) explained 4.59 % of the variance in the imitative
1352 compatibility effect ($F(3,181)=2.90$, $p=0.036$, $R^2=.04$, $f^2=0.05$) indicating a small effect.
1353 Mean RT was a predictor ($B=0.05$, $SEB=0.02$, $t(181)=2.79$, $p=.006$, $[0.01; 0.09]$), but both
1354 sex ($B=0.23$, $SEB=1.04$, $t(181)=0.22$, $p=.82$, $[-1.82; 2.29]$) and the sex*mean RT
1355 interaction ($B=-0.003$, $SEB=0.02$, $t(181)=-0.21$, $p=.83$, $[-0.04; 0.03]$) did not predict the
1356 imitative compatibility effect (see Figure 14).

1357 When the model included empathy, the model predicted 7.04 % of the variance.
1358 Empathic concern predicted the imitative compatibility effect above and beyond the
1359 base model ($B=-0.51$, $SEB=0.24$, $t(179)=-2.08$, $p=.04$, $[-0.99; -0.03]$), and explained an
1360 addition 2.3% of the variance ($\Delta R^2= .023$, $F(1,179)=4.35$, $p=.04$; Figure 15). A decrease
1361 in empathic concern predicted a higher imitative compatibility effect. When
1362 agreeableness and extraversion were included in the model, the model predicted 7.09%
1363 of the variance. Agreeableness marginally predicted the imitative compatibility effect
1364 ($B=-2.68$, $SEB=1.53$, $t(179)=-1.75$, $p=.081$, $[-5.67; 0.33]$) and explained an additional
1365 1.6% of the variance ($\Delta R^2= .016$, $F(1,179)=3.07$, $p=.08$). Extraversion also marginally
1366 predicted the imitative compatibility effect ($B=1.81$, $SEB=1.09$, $t(179)=1.67$, $p=.096$, $[-$
1367 $0.32; 3.96]$) and explained an additional 1.5% of the variance ($\Delta R^2= .015$, $F(1,179)=2.$
1368 78 , $p=.096$). Higher extraversion predicted higher imitative compatibility, whereas
1369 higher agreeableness predicted a lower imitative compatibility effect (see Figure 15).

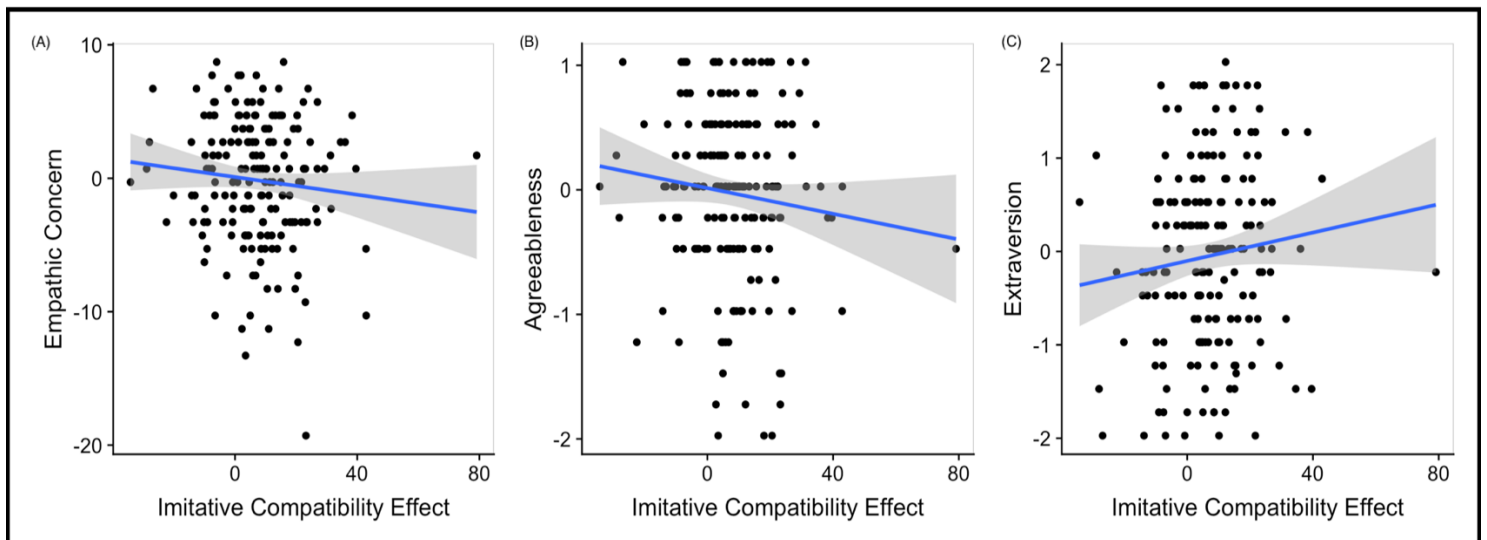
1370 Effect sizes attributable to the addition of empathy (Cohen's $f^2 = 0.03$), and
 1371 agreeableness and extraversion (Cohen's $f^2 = 0.03$) (beyond the base model) indicated
 1372 very small effects.



1373
 1374 **Figure 14.** *Experiment 3 – Multiple Regression Analyses.* Values of standardised coefficients are
 1375 plotted for each predictor variable (personality trait) along with their corresponding
 1376 uncertainties (95% confidence interval width for a normal distribution for each estimate).
 1377 Coefficients are standardised by dividing by two standard deviation units according to Gelman
 1378 (2008). The base model consists in the bottom three predictor variables (depicted in violet) –
 1379 mean RT, Sex, and meanRT*Sex. Abbreviations: RT = Reaction Time. *N.B.* The circles on the
 1380 purple distributions represent the standardised co-efficients for the main effect of meanRT,
 1381 main effect of sex, and meanRT*sex interaction respectively for each of the models tested. That
 1382 is why there are multiple circles for components of the base model because the base model was
 1383 part of all the models test (that is, one model for each personality trait).

1384

1385 Grandiose and vulnerability narcissism, autistic-like and schizotypal traits, and
 1386 alexithymia did not predict the imitative compatibility effect above and beyond the base
 1387 model (all p 's > 0.3, all CIs overlapping with zero; see Figure 14). The multiple
 1388 regression models are summarized in Supplementary Table 13. Zero-order correlations
 1389 are also consistent with the findings from the multiple regression analyses (see
 1390 Supplementary Table 14, Supplementary Figure 6).



1392 **Figure 15. Experiment 3 – Scatterplots.** Scatterplots depicting the relationship between imitative
 1393 compatibility effect and personality traits – empathic concern (A), agreeableness (B), and extraversion
 1394 (C). X axis denotes the imitative compatibility effect in milliseconds, and Y axis denotes mean centred
 1395 scores on the personality traits.

1396

1397 To evaluate the sex*trait interaction terms, we computed additional models –
 1398 each model consisted of the base model, one trait predictor (subscales were included in
 1399 the same model), and the sex*trait interaction term. None of the sex*trait interaction
 1400 terms predicted the compatibility effect above and beyond the base model
 1401 (Supplementary Figure 13). Multiple regression models are summarised in
 1402 Supplementary Table 15. Effect sizes attributable to the addition of the sex*trait
 1403 interaction terms (beyond the base model) indicated very small effects (Cohen's f^2 =

1404 <0.04 for all models). The pattern of results seen in Experiment 1 for the empathy*sex
1405 and alexithymia*sex models did not replicate in Experiment 3 (Figure 6C).

1406 Although our main question of interest was the link between personality traits
1407 and automatic imitation, for completeness, we also report results from the multiple
1408 regression analyses for spatial compatibility in the supplementary material (see
1409 Supplementary Tables 16 and 17, Supplementary Figures 14 and 15).

1410 *Stimuli rating.* All participants also rated the male and female hand stimuli on a
1411 scale of 1 to 9, with 1 being most masculine, and 9 being most feminine. All male hand
1412 stimuli were rated as masculine (Mean rating = 2.93, SD = 0.30). All female stimuli were
1413 rated as relatively feminine (Mean rating = 5.68, SD = 0.65). Although the female stimuli
1414 were not rated as strongly feminine, the ratings suggest that both male and female
1415 stimuli were perceived differently on average by the participants. The stimuli rating
1416 data is also available online.

1417 **Discussion**

1418 Results from Experiment 3 clearly show that a sex difference exists on the
1419 spatial compatibility effect such that females show a higher spatial compatibility effect
1420 than males. This difference did not persist when imitative compatibility was measured
1421 independently. This suggests that females and males do not differ in the control of
1422 automatic imitation as measured by the imitative compatibility effect.

1423 Furthermore, for the first time to date, we manipulated the sex of the stimuli
1424 across both male and female participants. Results indicated that there was no own-sex
1425 bias in the imitative compatibility effect. For the spatial compatibility effect, although
1426 the findings showed a trend toward an own-sex bias such that females showed a greater

1427 compatibility effect on female stimuli than male stimuli, this was a relatively small effect
1428 size, and thus does not explain much of the sex difference observed in the spatial
1429 compatibility effect.

1430 The findings from Experiment 3 thus suggest that it is unlikely that there is a
1431 sex difference in the imitative compatibility effect. Instead, our findings suggest that
1432 there is a sex difference in the spatial compatibility effect, which may reflect a difference
1433 in spatial control between males and females that in the case of this experiment is
1434 triggered by the location of a finger in space.

1435 The multiple regression analyses suggest that the imitative compatibility effect
1436 is invariant to stable traits of personality including grandiose and vulnerability
1437 narcissism, autistic-like and schizotypal traits, as well as alexithymia.

1438 Given prior evidence, we predicted that individuals who report higher empathy,
1439 extraversion, and agreeableness would be more prosocial, and would therefore imitate
1440 more than those who scored lower on these measures. In the current experiment,
1441 although higher extraversion predicted higher imitation, we found the opposite pattern
1442 for empathy and agreeableness. An increase in empathic concern and agreeableness
1443 predicted a decrease in the imitative compatibility effect. The effects, however, were
1444 small and predicted only an additional 2.45% (empathy) and 2.5% (extraversion and
1445 agreeableness) of the variance. Before making any firm conclusions, these results would
1446 need to be replicated using large sample sizes to ensure that these findings do not
1447 reflect false positives. In addition, none of the sex*trait interactions predicted the
1448 imitative compatibility effect, and the pattern of results from Experiment 1 for the
1449 sex*empathy and sex*alexithymia interactions did not replicate over Experiment 2 and
1450 3 (see Figure 6). Overall, therefore, these results provide only limited support for small

1451 or negligible effects of personality on automatic imitation reinforcing the suggestion
1452 that automatic imitation is largely invariant to stable traits of personality (Butler et al.,
1453 2015; Cracco et al., 2018).

1454 For all three experiments, we performed all the analyses again by further
1455 excluding participants who were three standard deviations away from the group mean
1456 on the compatibility effect on either of the tasks. For Experiment 1, no additional
1457 participants were excluded. For both Experiment 2 and 3, one additional participant
1458 was excluded. Obtained results were very similar to those reported above.

1459

1460

General Discussion

1461 By integrating methodological approaches from experimental and differential
1462 psychology, the current study shines new light on the relationships between stable
1463 features of individuals, such as personality and sex, and the architecture of cognitive
1464 control systems. Across three experiments, we consistently showed that cognitive
1465 control systems are largely invariant to stable aspects of personality, but exhibit a sex
1466 difference, such that females show greater interference than males. Moreover, we
1467 further qualified this sex difference in two ways. First, we showed that the sex
1468 difference was unrelated to the sex of the interaction partner and therefore did not
1469 reflect an in-group bias based on sex. Second, we showed that the sex difference was
1470 tied to a form of spatial interference control rather than imitative control and therefore
1471 it is unlikely to reflect a specialised mechanism for guiding social interactions
1472 exclusively. Instead, our findings suggest that a robust sex difference exists in the
1473 system (or set of subsystems) that operate in resolving a form of spatial interference

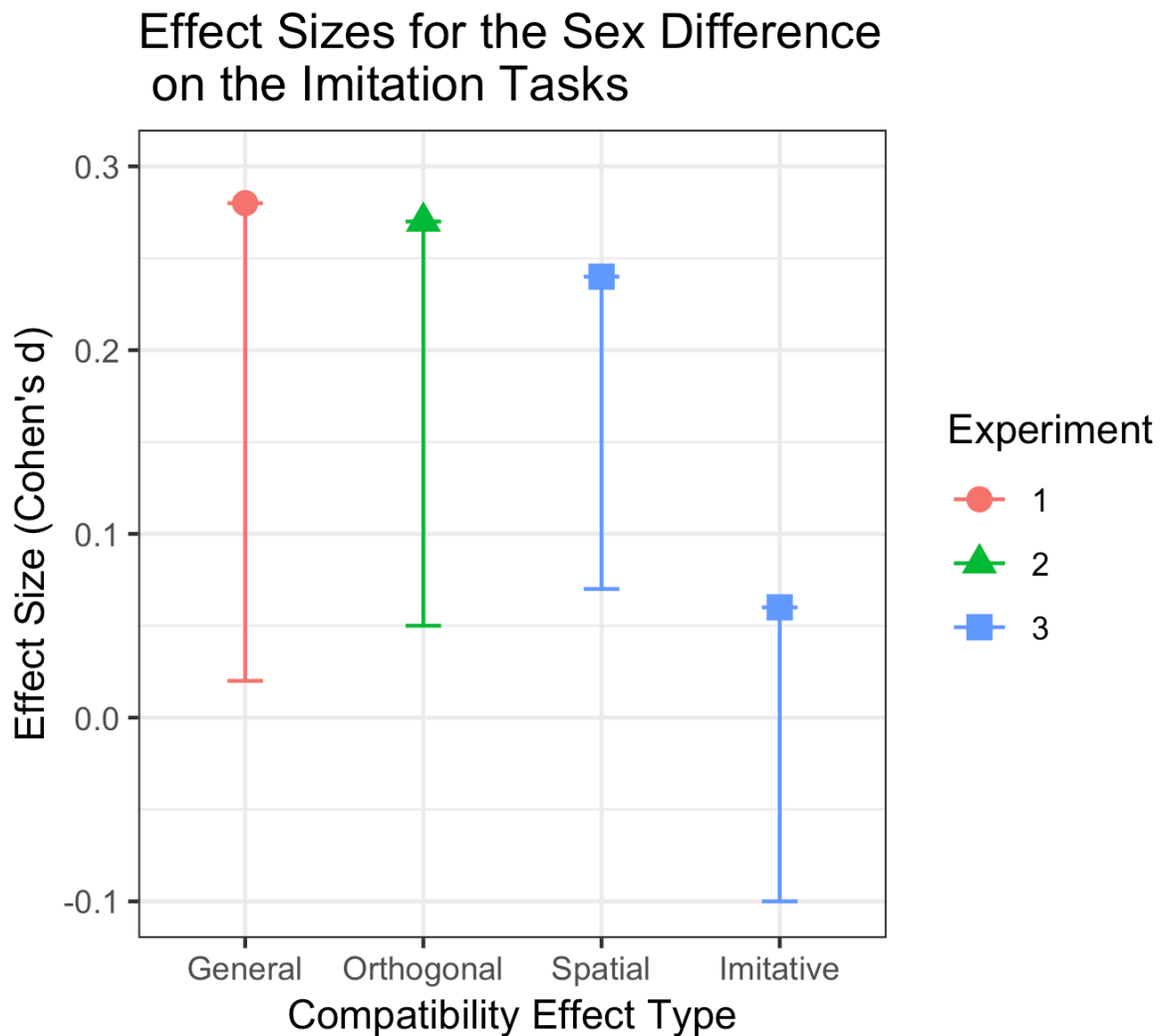
1474 control. The implications of these findings for understanding cognitive control systems
1475 in social and non-social contexts are discussed.

1476 **Are individual differences in interference control robust and replicable?**

1477 In recent years, a key question in psychology and neuroscience has concerned
1478 the credibility of reported findings (Open Science, 2015; Button et al., 2013; Pashler et
1479 al., 2012; Munafo et al., 2017; Vazire, 2018) with estimates of replicability ranging
1480 between 25 and 75% (Nosek & Lakens, 2014; Matzke et al., 2015; Marsman et al., 2017;
1481 Camerer et al., 2018). Studies that integrate experimental and differential approaches
1482 are rare in general, and in the context of imitation control, prior studies have typically
1483 used small sample sizes (Ainley et al., 2014; Chartrand & Bargh, 1999; Hogeveen & Obhi,
1484 2013; Obhi et al., 2013; Santiesteban et al., 2015). As such, one important contribution
1485 from the current study is a more robust and precise estimate of the size and
1486 replicability of sex differences in cognitive control. To do so, we used relatively large
1487 sample sizes, which could detect small-to-medium effect sizes with a high degree of
1488 confidence, and ran three separate experiments using designs that combined
1489 approaches from experimental and differential psychology.

1490 In Experiments 1 and 2, we replicated the sex difference found previously both
1491 when the SRC task measured automatic imitation as a composite of imitative and left-
1492 right spatial effects (Butler et al., 2015), as well as orthogonal spatial compatibility
1493 effects (Genschow et al., 2017). Furthermore, in Experiment 3, we measured imitative
1494 compatibility effects independent of spatial compatibility effects (Catmur & Heyes,
1495 2011; Berthenthal et al., 2006; Boyer et al., 2012; Jimenez et al., 2012). In Experiment 3,
1496 females showed a greater spatial compatibility effect than males, but there was no
1497 difference between the sexes on imitative compatibility. Thus, it is clear that the sex

1498 difference on the SRC task reflects a difference in spatial control between males and
1499 females, rather than a difference in a specialised system that is dedicated to social
1500 control.



1501 **Figure 16.** *Effect Sizes of the Sex Difference.* Cohen's d effect sizes of the sex difference on the
1502 imitation task (compatibility effect) across Experiments 1, 2, and 3. Error bar denotes one-tailed
1503 95% confidence interval.
1504

1505

1506 According to Cohen's benchmarks for interpreting effect sizes (Cohen, 1992), the
1507 difference between the sexes was a small-to-medium effect size (Cohen's $d = 0.28$) and
1508 was relatively consistent across the three experiments, with the lower bound of the
1509 95% CI > 0.02 (see Figure 16). Considering the sensitivity of our design, it is important

1510 to note that these effect sizes were below the 80% power mark, which our power
1511 analysis identified, as we had 80% power to detect effects greater than Cohen's d 0.36.
1512 Each individual experiment, therefore, has less than 80% power. This said, all three
1513 experiments showed results similar to Butler and colleagues (2015), in that they were
1514 in the same (predicted) direction and of a consistent magnitude even though the effect
1515 sizes were small. Further, consistent small effect sizes in large- N studies are more likely
1516 to represent the true state of nature than large effect sizes in small- N studies. In
1517 addition, these small effects can cumulate in their importance over time (Funder & Ozer,
1518 2019). By replicating the effects in separate large sample designs, it makes it less likely
1519 that these results represent sampling error (Zwaan et al., 2018). If we interpret the
1520 length of confidence interval (Armhein et al., 2019; Cumming, 2012), then our best
1521 estimate is a small to medium effect, with all likely effects being in the predicted
1522 direction (i.e., greater than zero). Therefore, building on prior work (Butler et al., 2015),
1523 across three large-sample experiments, we have provided a robust and relatively
1524 precise estimate of the size of the sex difference and shown that it reflects spatial rather
1525 than social control mechanisms.

1526 Moreover, across all three experiments, we consistently found that the control of
1527 automatic imitative tendencies, as measured by the SRC task, is invariant to differences
1528 in personality traits across individuals. Recently, it has been suggested that the
1529 investigation of an experimental effect at the group level, and individual differences
1530 within that effect are questions that can be at odds with each other. This is because
1531 group effects need low variability within the sample whereas differential psychology
1532 questions are dependent on high variability within the sample (Rogossa, 1988).
1533 Therefore, we cannot assume that robust experimental paradigms such as the SRC tasks
1534 used in the current experiments will lend themselves well to individual difference

1535 approaches. Thus, although a difference between groups can be detected if the groups
1536 means are sufficiently far away from each other to be detectable, the tasks used may not
1537 be able to distinguish between individuals in the population consistently (Hedge et al.,
1538 2017).

1539 Compared to prior studies (Obhi et al., 2013; Obhi & Hogeveen, 2013; Chartrand
1540 & Bargh, 1999), however, we provide a more robust test of hypotheses regarding
1541 individual differences as we used larger sample sizes, which produce higher statistical
1542 power, and we looked for consistent patterns of data across multiple experiments. By
1543 doing so, a more stable picture is emerging with regard to personality and SRC
1544 measures of automatic imitation, which suggests that mechanisms of imitative control
1545 are largely invariant to dimensions of personality (Butler et al., 2015; Genschow et al.,
1546 2017; Cracco et al., 2018), even when they are operating in more socially rich contexts
1547 (Exp. 1) and when spatial and imitative effects are more clearly separated (Exps. 2 and
1548 3). In short, any effects of personality were small and inconsistent across experiments.
1549 Of course, our design did not have sufficient power to detect small effects with
1550 reasonable confidence (> 80%), and such effects would require considerably larger
1551 sample sizes to be able to confidently confirm that they exist. Thus, our best estimate at
1552 present is that the effects of personality on SRC measures of automatic imitation are
1553 negligible or small.

1554 **Do differences in cognitive control reflect a sex difference or an in-group bias?**

1555 In Experiments 1 and 2, as well as in prior studies that have observed sex
1556 differences in the SRC imitation task (Butler et al., 2015; Genschow et al., 2017), the
1557 stimuli used were of a female hand. Thus, it was possible that the sex difference
1558 reflected an in-group bias leading to higher compatibility effects for females compared

1559 to males (Cracco et al., 2018). Indeed, there is already suggestive evidence (from studies
1560 with relatively small sample sizes), that both facial imitation and SRC measures of
1561 imitation have been found to increase when the interacting partner is an in-group
1562 member compared to an out-group member based on race, ethnicity, and arbitrary
1563 group assignment (Mondillon et al., 2007; Rauchbauer et al., 2015; Gleibs et al., 2016).

1564 In the current study, based on the sex of the interaction partner, we show no
1565 clear evidence for an own-sex bias for either spatial or imitative compatibility.

1566 Moreover, in terms of sensitivity, the use of a larger sample size than is typical means
1567 that our study had 80% power to detect effect sizes at or above Cohen's $d = 0.36$, which
1568 means that we can be reasonably confident that effect sizes of this magnitude or larger
1569 are unlikely. Taken together, although ingroup biases are potent in everyday life and
1570 relate to sex, race and ethnicity (Brown, 1995; Yee & Brown, 1995; Fischbein, 1996,
1571 Powlishta, 1995; Rudman & Goodwin, 2004; Kubota et al., 2012; Malpass & Kravitz,
1572 1969), the difference in interference control reported here reflects the sex of the
1573 participant, rather than an in-group bias based on the sex of the interaction partner. As
1574 such, these results are contrary to proposals put forward by Cracco and colleagues
1575 (2018), and highlight a stable individual difference in interference control, rather than
1576 an effect of the social context (i.e., the sex of the interaction partner).

1577 **What type of cognitive system underpins sex differences in interference control?**

1578 Three broad structures of cognitive system were candidates to underpin the sex
1579 difference in interference control: 1) a sex difference specific to social imitative control;
1580 2) a sex difference generalised across all types of control; 3) a sex difference specific to
1581 a form of non-social control. If the sex difference was solely tied to imitative control and
1582 reflected the workings of a specialised and domain-specific cognitive structure, we

1583 would have observed a sex difference only on the imitative compatibility component of
1584 the task. Likewise, if the sex difference reflected the operation of a straightforwardly
1585 domain-general system, we would have expected a difference between males and
1586 females on the flanker task, as well as both the spatial and imitative components of the
1587 automatic imitation SRC task. As such, these findings demonstrate that the sex
1588 difference is neither completely domain-general i.e., it does not generalise across all
1589 types of compatibility effects nor is it domain-specific i.e., it is not solely tied to the
1590 control of automatic imitation.

1591 Our findings show more support for the third type of cognitive system outlined
1592 above, which suggests that the sex difference reflects a particular type of non-social
1593 interference, which is not shared across all SRC tasks. Indeed, across our experiments,
1594 the sex difference was tied to a type of spatial interference observed in the spatial
1595 component of the automatic imitation SRC task, but not the imitative component of the
1596 same task or the non-social flanker task. A sex difference on spatial control, but not on
1597 imitative control, when measured on the same task, suggests that although general
1598 cognitive control systems are engaged for both tasks to some extent, they may not be
1599 engaged in an identical manner across both the compatibility effects. Moreover, it is
1600 unlikely that the sex difference on spatial compatibility reflects a difference in the
1601 perceptual processing of the social stimulus (i.e. the hand on the screen) as the stimuli
1602 are the same across both compatibility effects, but no sex difference emerges on the
1603 imitative compatibility effect. For both imitative and spatial compatibility, therefore, the
1604 input to the control mechanism that resolves conflict is the same i.e. a finger. However,
1605 the way conflict is resolved for spatial and imitative effects might involve mechanisms
1606 that operate differently as a function of sex.

1607 In addition to the sex difference not being tied to social or imitative control, it
1608 also reflects a component that is not shared with the flanker task. A lack of sex
1609 difference on the flanker task, and little or no correlation between the compatibility
1610 effects on the two tasks, has at least two possible interpretations, which are not
1611 mutually exclusive. First, it could reflect a lack of sensitivity. The differences between
1612 females and males on behavioural indices (such as RT) on the flanker task may be small
1613 (Clayson et al., 2011; Fischer et al., 2015; Stoet, 2011). In the current experiment, our
1614 sensitivity analysis suggests that we could detect effect sizes of Cohen's $d > 0.36$ with
1615 reasonable confidence (80%), but the effects of sex on the flanker were smaller than
1616 this in Experiment 1 and 2 (Cohen's $d = 0.15$ and 0.03 , respectively). Moreover, a large
1617 sample study with 895 participants found a small sex difference in the predicted
1618 direction on the flanker task using arrows such that females showed a greater
1619 compatibility effects than males (Fischer et al., 2015). Thus, there could also be a non-
1620 zero sex difference on the flanker task, but even if this turns out to be the case, it is clear
1621 that the size of the sex difference varies across different types of non-social cognitive
1622 control tasks.

1623 A second possible reason for the lack of sex difference on the flanker task is that
1624 the sex difference is underpinned by a particular type of non-social control. Previous
1625 studies help contextualise this finding by showing that females differ from males across
1626 a wide range of cognitive control tasks, especially those involving spatial processing
1627 (Bayliss & Tipper, 2005; Clayson et al., 2011; Stoet et al., 2017; Stoet, 2011). One
1628 possibility, therefore, is that the sex difference may reflect a difference in the two types
1629 of spatial conflict measured by the flanker and spatial compatibility effect. For example,
1630 in the SRC task measuring spatial compatibility, the conflict arises because a stimulus
1631 feature is inconsistent to the response, whereas the flanker task measures both

1632 stimulus-stimulus and stimulus-response conflicts (Kornblum et al., 1990; Kornblum,
1633 1994; Verbruggen et al., 2006). Further, it has been proposed that stimulus-response (S-
1634 R) conflicts and stimulus-stimulus (S-S) conflicts are underpinned by different
1635 processing patterns (Kornblum & Lee, 1995; Kornblum et al., 1999; Kornblum et al.,
1636 1990; Li et al., 2014; Frühholz et al., 2011; Zhang et al., 1999). Therefore, these types of
1637 conflicts would be worth taking into consideration in future research that investigates
1638 individual differences in automatic imitation, and social and non-social cognitive
1639 control.

1640 More generally, other sex differences, which do not rely on SRC paradigms, can
1641 further contextualise our findings. Indeed, prior research suggests that females differ
1642 from males on a range of social processes (Baron-Cohen, 2002). For example, females
1643 show greater empathy than males, which may lead to more pro-social behaviour, thus
1644 suggesting that females may imitate more than males (Christov-Moore et al., 2014;
1645 Baron-Cohen & Wheelwright, 2004; Schulte-Rüther et al., 2008). However, although
1646 empathy has been associated with a variety of paradigms investigating automatic
1647 imitation (Mueller et al., 2013; Sonnyby-Borgstrom 2002; Chartrand & Bargh, 1999),
1648 there does not seem to be a clear link between empathy and automatic imitation as
1649 measured on the SRC task (Butler et al., 2015; Genschow et al., 2017). Moreover, while
1650 females show higher facial mimicry than males (Sonnyby-Borgstrom 2002; 2008;
1651 Dimber, 1990; Hess & Bourgeois, 2010; Korb et al., 2015; Lundqvist, 1995), studies
1652 investigating imitation of other behaviours, such as nose-scratching, have not found any
1653 reliable sex differences, although such studies have been limited by small sample sizes
1654 (Chartrand & Bargh, 1999). Inconsistent and equivocal results across imitation tasks
1655 might suggest that these tasks engage different cognitive mechanisms.

1656 In addition, while we separated the spatial and imitative effects in Experiment 3
1657 using a modified version of the paradigm developed by Catmur and Heyes (2011), there
1658 are other tasks that measure imitative effects not confounded with spatial compatibility
1659 effects (e.g. Bortoletto et al., 2013). Thus, any conclusions we make our limited to the
1660 type of task we used in the current experiment. However, whether other tasks are more
1661 powerful measures of imitative effects is an empirical question that would need to be
1662 tested by future research.

1663 These findings raise a fundamental issue that remains unresolved does the SRC
1664 task actually link to imitation “in the wild” and social cognition more broadly. The
1665 domain of social cognition research has seen the use of cognitive psychology paradigms
1666 and methodologies to answer questions that are of interest to social psychologists
1667 (Lambert & Scherer, 2013). But are these paradigms truly measuring what we think
1668 they are measuring? More and more recent evidence suggests that social cognition is
1669 fundamentally different when we are involved in live social interactions with each other
1670 as compared to when we are doing tasks in a controlled environment (Schillbach et al.,
1671 2013; Redcay & Schillbach, 2019). For instance, researchers have suggested that eye-
1672 gaze behaviours when measured using screen-based tasks cannot be validly generalised
1673 to and used as a proxy for understanding gaze behaviours in live social interaction
1674 settings (Grossman, Zane, Mertens, & Mitchell, 2019). Thus, more empirical evidence is
1675 needed to know whether the SRC task of automatic imitation is actually measuring a
1676 social cognitive process (or just one component of a multi-dimensional construct like
1677 imitation).

1678 Indeed, although different measures of automatic imitation have been previously
1679 assumed to rely on the same underlying mechanisms, there is accumulating empirical
1680 and theoretical reason to question such an assumption (Ramsey, 2018; Genschow et al.,

1681 2017). Thus, divergent sex differences across measures of different dimensions of
1682 automatic imitative behaviour may reflect differences in cognitive mechanisms that
1683 underpin these tasks.

1684

1685 **Limitations/ Constraints on generality**

1686 In the current work, we make conclusions about social and non-social cognitive
1687 control on the basis of compatibility effects measured on flanker and imitation tasks.

1688 This makes sense because we had predictions specifically about individual differences
1689 in social and non-social cognitive control as measured by compatibility effects.

1690 However, there are many processes that contribute to such compatibility effects, and
1691 cognitive control is only one of them. Therefore, we cannot be sure that individual
1692 differences (or lack thereof) on these tasks are tied to individual differences in cognitive
1693 control (Musslick et al., 2019). Thus, an important and complementary direction for
1694 future research would be to disentangle the extent to which inter-individual differences
1695 reflect differences in cognitive control or other cognitive capacities. To do so, one may
1696 include a baseline condition for both the flanker and imitation tasks, which may help to
1697 separate differences in cognitive control from differences in other factors that also
1698 contribute to SRC tasks.

1699 A second potential limitation to the current work is the use of the flanker task as
1700 a measure of non-social cognitive control. Cognitive control has various components
1701 which can be measured by many different tasks (Rondeel et al., 2015). However, we use
1702 different non-social tasks across the three experiments (i.e. flanker in Experiments 1
1703 and 2, and spatial compatibility in Experiment 3) as a comparison with the automatic
1704 imitation task and find similar results irrespective of the type of task we used.

1705 Therefore, although every task has drawbacks, we feel that our key findings are

1706 relatively robust in the sense that they remain largely indifferent to changes in
1707 experimental design.

1708

1709 **Conclusion**

1710 The current findings provide a general insight into the relationship between individual
1711 differences and cognitive control systems in social and non-social contexts. Integrating
1712 experimental and differential psychology approaches, across three large sample
1713 experiments, we show that there is negligible or no evidence for a link between social
1714 control and stable personality traits. However, cognitive control systems vary as a
1715 function of biological sex, such that females show a greater interference than males.
1716 Further, this sex difference does not reflect an in-group bias based on the sex of the
1717 interacting partner, and is not tied specifically to social control but reflects differences
1718 in the cognitive systems that operate in resolving a form of spatial interference.
1719 Therefore, we show that the sex difference exists in the system (or set of subsystems)
1720 that operate in resolving a form of spatial interference control, and that such systems
1721 are unaffected by social factors such as facial expression or the sex of the interaction
1722 partner. More generally, the results highlight the value of integrating approaches from
1723 experimental and differential psychology, as well as using large sample sizes, in order to
1724 investigate the relationship between cognitive control architectures and stable traits of
1725 individuals, which few studies have achieved to date.

1726

1727

References

- 1728
- 1729 Aboud, F. E. (1988). *Children and prejudice*. Oxford: Basil Blackwell.
- 1730 Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annual review of*
1731 *psychology, 60*, 693-716.
- 1732 Ainley, V., Brass, M. & Tsakiris, M. (2014). Heartfelt imitation: High interoceptive awareness is
1733 linked to greater automatic imitation. *Neuropsychologia, 60*, 21–28.
- 1734 Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2018). Raincloud plots: a multi-
1735 platform tool for robust data visualization. *PeerJ Preprints, 6*, e27137v1.
- 1736 Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief “red flags” for autism screening:
1737 The short autism spectrum quotient and the short quantitative checklist in 1,000 cases
1738 and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry,*
1739 *51(2)*, 202-212.
- 1740 Allport, G. W. (1954). *The nature of prejudice*. Cambridge/Reading, MA: Addison-Wesley.
- 1741 Alwall, N., Johansson, D., & Hansen, S. (2010). The gender difference in gaze-cueing: Associations
1742 with empathizing and systemizing. *Personality and Individual Differences, 49(7)*, 729-
1743 732.
- 1744 Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism.
1745 *Journal of Research in Personality, 40*, 440-350. doi:10.1016/j.jrp.2005.03.002
- 1746 Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004).
1747 Neural signals for the detection of unintentional race bias. *Psychological Science, 15(2)*,
1748 88-93.
- 1749 Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical
1750 significance. *Nature, 567*, 305-307

- 1751 Ashton, M. C., Paunonen, S. V., Helmes, E., & Jackson, D. N. (1998). Kin altruism, reciprocal
1752 altruism, and the Big Five personality factors. *Evolution and Human Behavior*, 19(4),
1753 243-255.
- 1754 Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I.
1755 Item selection and cross-validation of the factor structure. *Journal of psychosomatic*
1756 *research*, 38(1), 23-32.
- 1757 Baldauf, D., & Desimone, R. (2014). Neural Mechanisms of Object-Based Attention. *Science*, 344,
1758 424–427.
- 1759 Banich MT. 2009. Executive function: the search for an integrated account. *Curr. Dir. Psychol.*
1760 *Sci.* 18:89–94
- 1761 Barrio, V. D., Aluja, A., & García, L. F. (2004). Relationship between empathy and the Big Five
1762 personality traits in a sample of Spanish adolescents. *Social Behavior and Personality:*
1763 *An International Journal*, 32, 677–681. doi:10.2224/sbp.2004.32.7.677
- 1764 Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-
1765 spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism,
1766 males and females, scientists and mathematicians. *Journal of Autism and Developmental*
1767 *Disorders*, 31, 5–17.
- 1768 Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults
1769 with Asperger syndrome or high functioning autism, and normal sex differences. *Journal*
1770 *of Autism and Developmental Disorders*, 34, 163–175.
- 1771 Bayliss, A. P., di Pellegrino, G., & Tipper, S. P. (2005). Sex differences in eye gaze and symbolic
1772 cueing of attention. *Quarterly Journal of Experimental Psychology*, 58A, 631–650.
- 1773 Bernieri FJ (1988) Coordinated movement and rapport in teacher-student interactions. *Journal*
1774 *of Nonverbal Behavior* 12: 120–138

1775

1776 Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social
1777 categorization is sufficient to elicit an own-group bias in face recognition. *Psychological*
1778 *Science*, 18(8), 706-712.

1779 Bertenthal, B. I., Longo, M. R., & Kosobud, A. (2006). Imitative response tendencies following
1780 observation of intransitive actions. *Journal of Experimental Psychology: Human*
1781 *Perception and Performance*, 32, 210

1782 Bien, N., Roebroek, A., Goebel, R., & Sack, A. T. (2009). The brain's intention to imitate: The
1783 neurobiology of intentional versus automatic imitation. *Cerebral Cortex*, 19, 2338–
1784 2351.

1785 Binney, R., & Ramsey, R. (2019). Social Semantics: The role of conceptual knowledge and
1786 cognitive control in a neurobiological model of the social brain.

1787 Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring
1788 and cognitive control. *Psychological review*, 108(3), 624.

1789 Botvinick, M. M., & Cohen, J. D. (2014). The computational and neural basis of cognitive control:
1790 charted territory and new frontiers. *Cognitive science*, 38(6), 1249-1285.

1791 Boyer, T. W., Longo, M. R., & Bertenthal, B. I. (2012). Is automatic imitation a specialized form of
1792 stimulus–response compatibility? Dissociating imitative and spatial compatibilities. *Acta*
1793 *Psychologica*, 139, 440–448

1794 Brass M, Bekkering H, Wohlschlagler A, Prinz W. (2000). Compatibility between observed and
1795 executed finger movements: comparing symbolic, spatial, and imitative cues. *Brain and*
1796 *Cognition*, 44(2), 124–143. pmid:11041986

1797 Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence
1798 problem? *Trends in Cognitive Sciences*, 9(10), 489-495. doi:10.1016/j.tics.2005.08.007

1799

1800 Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition.
1801 Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences,
1802 364, 2359–2367.

1803 Brass, M., Zysset, S., & von Cramon, D. Y. (2001). The inhibition of imitative response tendencies.
1804 Neuroimage, 14, 1416–1423.

1805 Brown, R. (1995). Prejudice. Its social psychology. Blackwell, Oxford, U.K.

1806 Butler, E. E., Ward, R., & Ramsey, R. (2015). Investigating the Relationship between Stable
1807 Personality Characteristics and Automatic Imitation. PLoS ONE, 10(6), e0129651.
1808 doi:10.1371/journal.pone.012965

1809 Button, K. S. et al. Power failure: why small sample size undermines the reliability of
1810 neuroscience. Nat. Rev. Neurosci. 14, 365–376 (2013).

1811 Cacioppo, J. T., Berntson, G. G., Sheridan, J. F., and McClintock, M. K. (2000). Multi-level
1812 integrative analyses of human behavior: social neuroscience and the complementing
1813 nature of social and biological approaches. Psychol. Bull. 126, 829–843

1814 Camerer, C. F., Dreber, A., Holzmeister, H., H, T., Huber, J., Johannesson, M., et al. (2018).
1815 Evaluating the replicability of social science experiments in Nature and Science between
1816 2010 and 2015. Nature Human Behavior., 2, 637–644.

1817 Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2001). Children's lay theories about
1818 ingroups and outgroups: Reconceptualizing research on prejudice. Personality and
1819 Social Psychology Review, 5(2), 118-128.

1820 Campbell, R., Elgar, K., Kuntsi, J., Akers, R., Terstegge, J., Coleman, M., & Skuse, D. (2002). The
1821 classification of “fear” from faces is associated with face recognition skill in females.
1822 Neuropsychologia, 40 (6), 575–584. doi:10.1016/S0028-3932(01)00164-6

- 1823 Carling, K. (2000). Resistant outlier rules and the non-Gaussian case. *Stat. Data Anal.* 33, 249–
1824 258.
- 1825 Catmur, C., & Heyes, C. (2011). Time course analyses confirm independence of imitative and
1826 spatial compatibility. *Journal of Experimental Psychology: Human Perception and*
1827 *Performance*, 37(2), 409-421
- 1828 Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. Guilford
1829 Press.
- 1830 Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and
1831 social interaction. *Journal of personality and Social Psychology*, 76(6), 893-910.
- 1832 Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P.
1833 F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral*
1834 *Reviews*, 46, 604–627.
- 1835 Clayson, P. E., Clawson, A., & Larson, M. J. (2011). Sex differences in electrophysiological indices
1836 of conflict monitoring. *Biological Psychology*, 87, 282–289.
- 1837 Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, 112, 155–159.
- 1838 Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and*
1839 *meta-analysis*. New York: Routledge.
- 1840 Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., ... & Brass, M. (2018).
1841 Automatic imitation: A meta-analysis. *Psychological Bulletin*, 144(5), 453.
- 1842 Crescentini, C., Mengotti, P., Grecucci, A., & Rumiati, R. I. (2011). The effect of observed biological
1843 and non biological movements on action imitation: an fMRI study. *Brain Research*, 1420,
1844 80-92. doi:10.1016/j.brainres.2011.08.077
- 1845 Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American*
1846 *psychologist*, 12(11), 671.

1847 Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American*
1848 *psychologist*, 30(2), 116.

1849 de Schotten, M. T., and Shallice, T. (2017). Identical, similar or different? is a single brain model
1850 sufficient? *PLoS One* 86, 172–175. doi: 10.1016/j.cortex.2016.12.002

1851 Daniel, D. B., Pelotte, M., & Lewis, J. (2000). Lack of sex differences on the Stroop Color-Word
1852 Test across three age groups. *Perceptual and motor skills*, 90(2), 483-484.

1853 Darda, K. M., Butler, E. E., & Ramsey, R. (2018). Functional Specificity and Sex Differences in the
1854 Neural Circuits Supporting the Inhibition of Automatic Imitation. *Journal of cognitive*
1855 *neuroscience*, 30(6), 914-933.

1856 Darda, K. M., & Ramsey, R. (2019). The inhibition of automatic imitation: a meta-analysis and
1857 synthesis of fMRI studies. *NeuroImage*. 197:320-329.

1858 Davis, M. H. (1980). Interpersonal reactivity index (IRI). A multidimensional approach to
1859 individual differences in empathy. *JSAS Catalog of Selected Documents in*
1860 *Psychology*, 10, 85.

1861 Dimberg U (1982) Facial reactions to facial expressions. *Psychophysiology* 19: 643–647.

1862 Dodell-Feder, D., Ressler, K. J., & Germine, L. T. (2019). Social cognition or social class and
1863 culture? On the interpretation of differences in social cognitive
1864 performance. *Psychological medicine*, 1-13.

1865 Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for
1866 intelligent behaviour. *Trends in Cognitive Sciences*, 14, 172–179.

1867 Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-
1868 effective measures of the big five factors of personality. *Psychological Assessment*,
1869 18(2), 192-203. DOI: 10.1037/1040-3590.18.2.192

1870 Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target
1871 letter in a nonsearch task. *Perception and Psychophysics*, 16, 143–149.

1872 Eysenck, H. J. (1997). Personality and experimental psychology: the unification of psychology
1873 and the possibility of a paradigm. *Journal of Personality and social Psychology*, 73(6),
1874 1224.

1875 Eysenck, H. J., & Eysenck, M. W. (1985). Personality and individual differences: a natural science
1876 approach. Plenum Press. New York.

1877 Fischer, A. G., Danielmeier, C., Villringer, A., Klein, T. A., & Ullsperger, M. (2016). Gender
1878 influences on brain responses to errors and post-error adjustments. *Scientific reports*, 6,
1879 24435.

1880 Fischer-Baum, S., Kook, J. H., Lee, Y., Ramos Nuñez, A. I., & Vannucci, M. (2018). Individual
1881 Differences in the Neural and Cognitive Mechanisms of Single Word Reading. *Frontiers*
1882 *in human neuroscience*, 12, 271.

1883 Fishbein H. D. 1996. *Peer Prejudice and Discrimination: Evolutionary, Cultural, and Develop-*
1884 *mental Dynamics*. Westview Press: Boulder, CO

1885 Frith, C. D. (2008). Social cognition. *Philosophical Transactions of the Royal Society B: Biological*
1886 *Sciences*, 363(1499), 2033-2039.

1887 Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual review of psychology*, 63,
1888 287-313.

1889 Frühholz, S., Godde, B., Finke, M., & Herrmann, M. (2011). Spatio-temporal brain dynamics in a
1890 combined stimulus–stimulus and stimulus–response conflict task. *Neuroimage*, 54(1),
1891 622-634.

1892 Geary, D. C. (2010). *Male, female: The evolution of human sex differences*. American
1893 Psychological Association.

- 1894 Genschow, O., van Den Bossche, S., Cracco, E., Bardi, L., Rigoni, D., & Brass, M. (2017). Mimicry
1895 and automatic imitation are not correlated. *PLoS One*, 12, e0183784.
- 1896 Gigerenzer, G. (2018). Statistical Rituals: The Replication Delusion and How We Got There.
1897 *Advances in Methods and Practices in Psychological Science*, 2515245918771329.
- 1898 Gowen, E., Bolton, E., & Poliakoff, E. (2016). Believe it or not: Moving non-biological stimuli
1899 believed to have human origin can be represented as human movement. *Cognition*,
1900 146, 431–438.
- 1901 Gleibs, I. H., Wilson, N., Reddy, G., & Catmur, C. (2016). Group dynamics in automatic
1902 imitation. *PloS one*, 11(9), e0162880.
- 1903 Graziano, W. G., & Eisenberg, N. (1997). Agreeableness: A dimension of personality.
1904 In *Handbook of personality psychology* (pp. 795-824). Academic Press.
- 1905 Grecucci, A., Koch, I., & Rumiati, R. I. (2011). The role of emotional context in facilitating
1906 imitative actions. *Acta Psychologica*, 138(2), 311e315.
1907 <http://dx.doi.org/10.1016/j.actpsy.2011.07.005>.
- 1908 Grossman, R. B., Zane, E., Mertens, J., & Mitchell, T. (2019). facetime vs. Screentime: Gaze
1909 patterns to Live and Video Social Stimuli in Adolescents with ASD. *Scientific*
1910 *reports*, 9(1), 1-10.
- 1911 Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological bulletin*, 85(4), 845.
- 1912 Hamilton, A. F. de C. (2013). Reflecting on the mirror neuron system in autism: A systematic
1913 review of current theories. *Developmental Cognitive Neuroscience*, 3, 91–105. doi:
1914 10.1016/j.dcn.2012.09.008
- 1915 Hansen J, Alves H, Trope Y (2016) Psychological distance reduces literal imitation: Evidence
1916 from an imitation-learning paradigm. *Journal of Experimental Psychology: Human*
1917 *Perception and Performance*: 320–330. doi: 10.1037/xhp0000150

1918 Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of
1919 Murray's Narcism Scale. *Journal of research in personality*, 31(4), 588-599.

1920 Hess, U., & Bourgeois, P. (2010). You smile–I smile: Emotion expression in social
1921 interaction. *Biological psychology*, 84(3), 514-520

1922 Heyes, C. (2009). Evolution, development and intentional control of imitation. *Philosophical*
1923 *Transactions of the Royal Society of London. Series B: Biological Sciences*, 364(1528),
1924 2293–2298. doi:10.1098/rstb. 2009.0049

1925 Heyes, C. (2011). Automatic imitation. *Psychological bulletin*, 137(3), 463.

1926 Hogeveen, J., & Obhi, S. S. (2013). Automatic imitation is automatic, but less so for narcissists.
1927 *Experimental Brain Research*, 224, 613-621. doi: 10.1007/s00221-012-3339-6

1928 Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, 65, 373–
1929 398. Crossref, Google Scholar

1930 Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive
1931 control. *Trends in cognitive sciences*, 19(3), 126-132.

1932 Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in cognitive*
1933 *sciences*, 13(12), 524-531.

1934 JASP Team (2018). JASP (Version 0.9)[Computer software]

1935 Jeffreys, H. (1961). *Theory of Probability*. Oxford: UK: Oxford University Press.

1936 Jiménez, L., Recio, S., Méndez, A., Lorda, M. J., Permuy, B., & Méndez, C. (2012). Automatic
1937 imitation and spatial compatibility in a key-pressing task. *Acta psychologica*, 141(1), 96-
1938 103.

1939 Judge, J., & Taylor, P. J. (2012). Gender differences on the semantic flanker task using
1940 transposed-letter target words. *The Quarterly Journal of Experimental*
1941 *Psychology*, 65(10), 2008-2017.

- 1942 Kao, L. S., & Green, C. E. (2008). Analysis of variance: is there a difference in means and what
 1943 does it mean?. *Journal of Surgical Research*, 144(1), 158-170.
- 1944 Kavanagh, L. C., & Winkielman, P. (2016). The functionality of spontaneous mimicry and its
 1945 influences on affiliation: An implicit socialization account. *Frontiers in psychology*, 7,
 1946 458.
- 1947 Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological
 1948 movement on action. *Curr Biol*, 13(6), 522-525.
- 1949 Korb, S., Malsert, J., Rochas, V., Rihs, T. A., Rieger, S. W., Schwab, S., ... & Grandjean, D. (2015).
 1950 Gender differences in the neural network of facial mimicry of smiles—An rTMS
 1951 study. *Cortex*, 70, 101-114.
- 1952 Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for
 1953 stimulus-response compatibility--a model and taxonomy. *Psychological review*, 97(2),
 1954 253.
- 1955 Kornblum, S., & Lee, J. W. (1995). Stimulus-response compatibility with relevant and irrelevant
 1956 stimulus dimensions that do and do not overlap with the response. *Journal of*
 1957 *Experimental Psychology: Human Perception and Performance*, 21(4), 855.
- 1958 Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they
 1959 overlap with: The case of Stroop- and Simon-like stimuli. *Psychological*
 1960 *Research/Psychologische For-schung*, 56, 130-135.
- 1961 Kornblum, S., Stevens, G. T., Whipple, A., & Requin, J. (1999). The effects of irrelevant stimuli: 1.
 1962 The time course of stimulus-stimulus and stimulus-response consistency effects with
 1963 Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of*
 1964 *Experimental Psychology: Human Perception and Performance*, 25(3), 688.

- 1965 Krach, S., Blumel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., (2009). Are women
1966 better mindreaders? Sex differences in neural correlates of mentalizing detected with
1967 functional MRI. *BMC Neuroscience*, 10, 9.
- 1968 Kubota, J. T., Banaji, M. R., & Phelps, E. A. (2012). The neuroscience of race. *Nature*
1969 *neuroscience*, 15(7), 940.
- 1970 Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A
1971 practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.
- 1972 Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create
1973 affiliation and rapport. *Psychological Science*, 14(4), 334e339.
1974 <http://dx.doi.org/10.1111/1467-9280.14481>.
- 1975 Lambert, A. J., & Scherer, L. (2013). Measurement and methodology in social cognition: A
1976 historical perspective. *The Oxford Handbook of Social Cognition*, 33.
- 1977 Larsen, H., Overbeek, G., Granic, I., & Engels, R. C. M. E. (2010). Imitation of alcohol consumption
1978 in same-sex and other-sex dyads. *Alcohol and Alcoholism*, 45(6), 557-562. doi:
1979 10.1093/11calc/agq053
- 1980 Levant, R. F., Hall, R. J., Williams, C. M., & Hasan, N. T. (2009). Gender differences in
1981 alexithymia. *Psychology of men & masculinity*, 10(3), 190.
- 1982 Li, Q., Nan, W., Wang, K., & Liu, X. (2014). Independent processing of stimulus-stimulus and
1983 stimulus-response conflicts. *PLoS One*, 9(2), e89249.
- 1984 Lundqvist, L. O. (1995). Facial EMG reactions to facial expressions: A case of facial emotional
1985 contagion? *Scandinavian journal of psychology*, 36(2), 130-141.
- 1986 MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative
1987 review. *Psychological bulletin*, 109(2), 163.

- 1988 Malpass, R.S., & Kravitz, J. (1969). Recognition for faces of own and other "race." *Journal of*
 1989 *Personality and Social Psychology*, 13, 330–334
- 1990 Marsh, L. E., Bird, G., & Catmur, C. (2016). The imitation game: Effects of social cues on
 1991 'imitation' are domain-general in nature. *Neuroimage*, 139, 368–375.
- 1992 Marsman, M., Schönbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., & Wagenmakers, E. J. (2017). A
 1993 Bayesian bird's eye view of 'Replications of important results in social
 1994 psychology'. *Royal Society Open Science*, 4(1), 160426.
- 1995 Matzke, D., Nieuwenhuis, S., van Rijn, H., Slagter, H. A., van der Molen, M. W. & Wagenmakers, E.-
 1996 J. (2015) The effect of horizontal eye movements on free recall: A preregistered
 1997 adversarial collaboration. *Journal of Experimental Psychology: General* 144(1):e1–15.
- 1998 McCrae, R. R., & Costa Jr, P. T. (1999). A five-factor theory of personality. *Handbook of*
 1999 *personality: Theory and research*, 2(1999), 139-153.
- 2000 Merritt, P., Hirshman, E., Wharton, W., Stangl, B., Devlin, J., & Lenz, A. (2007). Evidence for
 2001 gender differences in visual selective attention. *Personality and individual*
 2002 *differences*, 43(3), 597-609.
- 2003 Michael, J., & D'Ausilio, A. (2015). Domain-specific and domain-general processes in social
 2004 perception—A complementary approach. *Consciousness and cognition*, 36, 434-437.
- 2005 Miller DI, Halpern DF (2014) The new science of cognitive sex differences. *Trends Cogn Sci*
 2006 18(1):37–45. doi:10.1016/j.tics.2013.10.011
- 2007 Milner D. 1983. *Children and Race: Ten Years On*. Ward Lock Educational: London. Nesdale D.
 2008 2000. Developmental changes in children's ethnic preferences and social
 2009 cognition. *Journal of Applied Developmental Psychology* 20: 501–519.

- 2010 Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The
2011 unity and diversity of executive functions and their contributions to complex “frontal
2012 lobe” tasks: A latent variable analysis. *Cognitive psychology*, 41(1), 49-100.
- 2013 Mondillon, L., Niedenthal, P. M., Gil, S., & Droit-Volet, S. (2007). Imitation of in-group versus
2014 outgroup members’ facial expressions of anger: a test with a time perception task. *Social
2015 Neuroscience*, 2(3-4), 223–37. <http://doi.org/10.1080/17470910701376894>
- 2016 Müller, B. C., Leeuwen, M. L., Baaren, R. B., Bekkering, H., & Dijksterhuis, A. (2013). Empathy is a
2017 beautiful thing: Empathy predicts imitation only for attractive others. *Scandinavian
2018 Journal of Psychology*, 54, 401–406.
- 2019 Musslick, S., Cohen, J. D., & Shenhav, A. (2019). Decomposing individual differences in cognitive
2020 control: A model-based approach. In *Proceedings of the 41st Annual Meeting of the
2021 Cognitive Science Society* (pp. 2427-2433). Cognitive Science Society Montreal, CA.
- 2022 Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., . . .
2023 Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human
2024 Behaviour*, 1, 0021. doi:10.1038/s41562-016-0021
- 2025 Nosek, B. A., & Lakens, D. (2014). Registered reports.
- 2026 Oberman, L. M., & Ramachandran, V. S. (2007). The simulating social mind: the role of the mirror
2027 neuron system and simulation in the social and communicative deficits of autism
2028 spectrum disorders. *Psychological Bulletin*, 133(2), 310-327. doi: 10.1037/0033-
2029 2909.133.2.310
- 2030 Obhi, S. S., Hogeveen, J., Giacomin, M., & Jordan, C. H. (2013). Automatic imitation is reduced in
2031 narcissists. *Journal of Experimental Psychology: Human Perception and Performance*,
2032 40(3), 920-928. doi: 10.1037/a0034056
- 2033 Ochsner, K. N., & Lieberman, M. D. (2001). The emergence of social cognitive
2034 neuroscience. *American Psychologist*, 56(9), 717.

2035 Open Science Collaboration. (2015). Estimating the reproducibility of psychological
2036 science. *Science*, 349, 6251.

2037 O'Reilly RC, Herd SA, Pauli WM. 2010. Computational models of cognitive control. *Curr. Opin.*
2038 *Neurobiol.* 20: 257–61

2039 Pashler, H., Coburn, N., Harris, C. R. (2012). Priming of Social Distance? Failure to Replicate
2040 Effects on Social and Food Judgments. *PLoS ONE*, 7(8), e42510.
2041 doi:10.1371/journal.pone.0042510.

2042 Payne, B. K. (2005). Conceptualizing control in social cognition: How executive functioning
2043 modulates the expression of automatic stereotyping. *Journal of personality and social*
2044 *psychology*, 89(4), 488.

2045 Pernet, C.R., Wilcox, R. & Rousselet, G.A. (2013). Robust correlation analyses: false positive and
2046 power validation using a new open source Matlab toolbox. *Front. in Psychology*, 3, 606.
2047 doi: 10.3389/fpsyg.2012.00606

2048 Powlishta K. K. 1995. Intergroup processes in childhood: Social categorization and sex role
2049 development. *Developmental Psychology*31: 781–788.

2050 Ramsey, R. (2018). What are reaction time indices of automatic imitation
2051 measuring?. *Consciousness and cognition*, 65, 240-254.

2052 Rauchbauer, B., Majdandžić, J., Hummer, A., Windischberger, C., & Lamm, C. (2015). Distinct
2053 neural processes are engaged in the modulation of mimicry by social group-membership
2054 and emotional expressions. *Cortex*, 70, 49-67, doi:10.1016/j.cortex.2015.03.007

2055 Ray, E., & Heyes, C. (2011). Imitation in infancy: the wealth of the stimulus. *Developmental*
2056 *science*, 14(1), 92-105.

- 2057 Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Sex, sexual orientation, and identification of
2058 positive and negative facial affect. *Brain and Cognition*, 54, 179-185.
2059 doi:10.1016/j.bandc.2004.01.002
- 2060 Raine, A., & Benishay, D. (1995). The SPQ-B: A brief screening instrument for schizotypal
2061 personality disorder. *Journal of Personality Disorders*, 9(4), 346-355.
- 2062 Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the
2063 mechanisms of social interaction. *Nature Reviews Neuroscience*, 20(8), 495-505.
- 2064 Rondeel, E., Van Steenbergen, H., Holland, R., & van Knippenberg, A. (2015). A closer look at
2065 cognitive control: Differences in resource allocation during updating, inhibition and
2066 switching as revealed by pupillometry. *Frontiers in human neuroscience*, 9, 494.
- 2067 Rousseeuw, P.J. (1984), "Least Median of Squares Regression," *Journal of the American*
2068 *Statistical Association*, Vol. 79, pp. 871-881.
- 2069 Rousseeuw, P.J. and Van Driessen, K. (1999), "A Fast Algorithm for the Minimum Covariance
2070 Determinant Estimator," *Technometrics*, 41, pp. 212-223.
- 2071 Rubia, K., Hyde, Z., Halari, R., Giampietro, V., & Smith, A. (2010). Effects of age and sex on
2072 developmental neural networks of visual-spatial attention allocation. *Neuroimage*,
2073 51, 817-827.
- 2074 Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do
2075 women like women more than men like men?. *Journal of personality and social*
2076 *psychology*, 87(4), 494.
- 2077 Russell, T. A., Tchanturia, K., Rahman, Q., & Schmidt, U. (2007). Sex differences in theory of mind:
2078 A male advantage on Happé's "cartoon" task. *Cognition and Emotion*, 21, 1554-1564.
2079 doi: 10.1080/02699930601117096

2080 Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., & Bird, G. (2012). Training social
2081 cognition: from imitation to theory of mind. *Cognition*, 122(2), 228-235.

2082 Santiesteban, I., Bird, G., Tew, O., Cioffi, M. C., & Banissy, M. J. (2015). Mirror-touch synaesthesia:
2083 Difficulties inhibiting the other. *Cortex*, 71, 116-121, doi:10.1016/j.cortex.2015.06.019

2084 Schilbach L., Eickhoff S. B., Cieslik E. C., Kuzmanovic B., Vogeley K. (2011). Shall we do this
2085 together Social gaze influences action control in a comparison group, but not in
2086 individuals with high-functioning autism. *Autism* 16, 151-
2087 16210.1177/1362361311409258

2088 Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013).
2089 Toward a second-person neuroscience 1. *Behavioral and brain sciences*, 36(4), 393-414.

2090 Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in*
2091 *Cognitive Science*, 1(2), 353-367.

2092 Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman?
2093 Sex differences in big five personality traits across 55 cultures. *Journal of Personality*
2094 *and Social Psychology*, 94(1), 168-182. doi: 10.1037/0022-3514.94.1.168

2095 Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender
2096 differences in brain networks supporting empathy. *Neuroimage*, 42(1), 393-403.

2097 Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's
2098 preferences for novel objects. *Developmental science*, 13(4), 599-610.

2099 Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed
2100 flexibility in data collection and analysis allows presenting anything as
2101 significant. *Psychological science*, 22(11), 1359-1366.

2102 Sonnby-Borgström, M. (2002). Automatic mimicry reactions as related to differences in
2103 emotional empathy. *Scandinavian Journal of Psychology*, 43, 433-443.

- 2104 Sonnby-Borgström, M., Jönsson, P., & Svensson, O. (2008). Gender differences in facial imitation
2105 and verbally reported emotional contagion from spontaneous to emotionally regulated
2106 processing levels. *Scandinavian Journal of Psychology*, 49, 111–122. doi:
2107 10.1111/j.1467-9450.2008.00626.x
- 2108 Southgate, V., & Hamilton, A. F. de C. (2008). Unbroken mirrors: challenging a theory of Autism.
2109 *Trends in Cognitive Sciences*, 12(6), 225-229. doi: 10.1016/j.tics.2008.03.005
- 2110 Sowden, S., Brewer, R., Catmur, C., & Bird, G. (2016). The specificity of the link between
2111 alexithymia, interoception, and imitation. *Journal of Experimental Psychology: Human*
2112 *Perception and Performance*, 42(11), 1687.
- 2113 Spunt, R. P., & Adolphs, R. (2017). A new look at domain specificity: Insights from social
2114 neuroscience. *Nature Reviews Neuroscience*, 18, 559.
- 2115 Stoet, G. (2010). Sex differences in the processing of flankers. *The Quarterly Journal of*
2116 *Experimental Psychology*, 63, 633–638.
- 2117 Stoet, G. (2017). Sex differences in the Simon task help to interpret sex differences in selective
2118 attention. *Psychological Research*, 81, 571–581.
- 2119 Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures
2120 and postures: A study of imitation. *Journal of Experimental Psychology: Human*
2121 *Perception and Performance*, 26, 1746–1759. doi:10.1037/0096-1523.26.6.1746
- 2122 Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup
2123 behaviour. *European journal of social psychology*, 1(2), 149-178.
- 2124 Thakkar, K. N., Peterman, J. S., & Park, S. (2014). Altered brain activation during action imitation
2125 and observation in schizophrenia: A translational approach to investigating social
2126 dysfunction in schizophrenia. *American Journal of Psychiatry*, 171(5), 539-548.

- 2127 Trujillo-Ortiz, A., R. Hernandez-Walls, K. Barba-Rojo and L. Cupul-Magana. (2007).
2128 HZmvntest:Henze-Zirkler's Multivariate Normality Test. A MATLAB file.
- 2129 Van Baaren, R. B., Holland, R. W., Steenaert, B., & van Knippenberg, A. (2003). Mimicry for
2130 money: Behavioral consequences of imitation. *Journal of Experimental Social*
2131 *Psychology*, 39(4), 393-398.
- 2132 van Baaren, R., Janssen, L., Chartrand, T. L., & Dijksterhuis, A. (2009). Where is the love? The
2133 social aspects of mimicry. *Philosophical Transactions of the Royal Society B: Biological*
2134 *Sciences*, 364(1528), 2381-2389. doi:10.1098/rstb.2009.0057
- 2135 Van Bavel, J. J., & Cunningham, W. A. (2009). Self-categorization with a novel mixed-race group
2136 moderates automatic social and racial biases. *Personality and Social Psychology*
2137 *Bulletin*, 35(3), 321-335.
- 2138 Van der Elst, W., Van Boxtel, M. P. J., Van Breakelen, G. J. P., Jolles, J. (2006). The Stroop Color-
2139 Word Test: Influence of Age, Sex, and Education; and Normative Data for a Large Sample
2140 Across the Adult Age Range. *Assessment*, 13(1), 62-79, doi:
2141 10.1177/1073191105283427
- 2142 Van der Graaff, J., Branje, S., De Wied, M., Hawk, S., Van Lier, P., & Meeus, W. (2014). Perspective
2143 taking and empathic concern in adolescence: Gender differences in developmental
2144 changes. *Developmental psychology*, 50(3), 881.
- 2145 Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and
2146 progress. *Perspectives on Psychological Science*, 13(4), 411-417.
- 2147 Verboten, S., & Hubert, M. (2005). LIBRA: a MATLAB Library for Robust Analysis, Chemometrics
2148 and Intelligent Laboratory Systems(75), 127-136.
- 2149 Verbruggen, F., Notebaert, W., Liefoghe, B., & Vandierendonck, A. (2006). Stimulus-and
2150 response-conflict-induced cognitive control in the flanker task. *Psychonomic Bulletin &*
2151 *Review*, 13(2), 328-333.

2152 Wang, Y., Ramsey, R., & Hamilton, A. F. D. C. (2011). The control of mimicry by eye contact is
2153 mediated by medial prefrontal cortex. *Journal of Neuroscience*, 31(33), 12001-12010.

2154 Wang, Y., & Hamilton, A. F. de C. (2014). Why does gaze enhance mimicry? Placing gaze-mimicry
2155 effects in relation to other gaze phenomena. *Quarterly Journal of Experimental*
2156 *Psychology*, 67(4), 747-762. doi:10.1080/17470218.2013.828316

2157 Webb JT (1972) Interview synchrony: An investigation of two speech rate measures in an
2158 automated standardized interview In: Pope B, Siegman AW, editors. *Studies in dyadic*
2159 *communication*. New York: Pergamon; pp. 115–133.

2160 Wilcox, R. R. (2012a). *Introduction to Robust Estimation and Hypothesis Testing*, 3rd Edn.
2161 Oxford: Academic Press.

2162 Williams, J. H., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and
2163 autism. *Neuroscience and Biobehaviour Review*, 25, 287–295. doi: 10.1016/S0149-
2164 7634(01)00014-8

2165 Yee M, Brown R. 1994. The development of gender differentiation in young children. *British*
2166 *Journal of Social Psychology*33: 183–196.

2167 Zaki, J., Hennigan, K., Weber, J., & Ochsner, K. N. (2010). Social cognitive conflict resolution:
2168 Contributions of domain general and domain specific neural systems. *Journal of*
2169 *Neuroscience*, 30, 8481–8488.

2170 Zhang, H. H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of
2171 stimulus–stimulus and stimulus–response compatibility. *Cognitive psychology*, 38(3),
2172 386-432.

2173 Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication
2174 mainstream. *Behavioral and Brain Sciences*, 41.

2175

2176

2177