Psicothema 2009. Vol. 21, nº 3, pp. 382-389
www.psicothema.com
ISSN 0214 - 9915 CODEN PSOTEG

Replicated nil associations of digit ratio (2D:4D) and absolute finger lengths with implicit and explicit measures of aggression

Martin Voracek and Stefan Stieger
University of Vienna (Austria)

Preliminary evidence suggests that within-sex individual variation in the length ratio of the index finger to the ring finger (2D:4D), a putative marker of prenatal androgen levels, may be more strongly or consistently related to implicit measures (Implicit Association Test, IAT) than to corresponding explicit (self-report) measures of candidate personality traits. The underlying logic is that the non-introspective IAT may reflect earlier and inaccessible experiences, whereas introspective self-reports reflect more recent and necessarily accessible experiences. Associations of 2D:4D and absolute finger length (a marker of pubertal-adolescent androgen levels) with implicit versus explicit aggression measures (IAT vs. Buss-Perry Aggression Questionnaire and a feeling thermometer) were examined in two samples with identical procedures (*Ns*= 244 and 233). Attesting to procedural validity and data typicality, several experiments of related research were replicated in both samples, including sex differences in 2D:4D and theory compliant interrelations among psychometric measures. However, no theory compliant, reliable, or replicable associations of 2D:4D or finger length with implicit or explicit aggression measures resulted. These nil findings cast doubt on hypothesized advantages of implicitly (over explicitly) measured target traits for 2D:4D research. More generally, they add to a growing number of replication failures in this area.

Replicación de asociaciones nulas del ratio del segundo y cuarto dedo (2D:4D) y la longitud absoluta de los dedos con medidas de agresión implícitas y explícitas. Evidencias preliminares sugieren que la variabilidad intersexual en la relación entre la longitud del 2º dedo (índice) y del cuarto dedo (2D:4D), como marcador indirecto de los niveles de andrógenos prenatales, puede poseer una fuerte y coherente relación con medidas implícitas (Test de Asociación Implícita, IAT) que corresponden a medidas explícitas (autoinforme) de rasgos de la personalidad. La lógica fundamental es que el IAT-no introspectivo puede reflejar las experiencias más tempranas e inaccesibles, mientras que los autoinformes reflejan las experiencias más recientes y necesariamente accesibles. La asociación entre el 2D:4D y la longitud absoluta del dedo (como marcador de los niveles de andrógenos en la etapa de la pubertad-adolescencia) con las medidas de agresión implícitas vs las explícitas (IAT vs. Cuestionario de Agresión de Buss-Perry y termómetro de sentimiento) fueron analizadas mediante los mismos procedimientos en dos muestras (*Ns*= 244 and 233). Para verificar la validez del procedimiento y la normalidad de los datos, varias experiencias de la investigación fueron replicadas en ambas muestras, incluyendo las diferencias sexuales en el 2D:4D y las de la interrelación de la teoría conformista entre medidas psicométricas.

The rediscovery of sex differences (men < women) in the length ratio of the index finger (2D) to the ring finger (4D) (2D:4D; Manning, Scutt, Wilson, & Lewis-Jones, 1998), along with a hypothesis about the origin of these, has spurred ongoing research interest in psychology and adjacent disciplines. Because this sex effect is determined in utero and 2D:4D shows little plasticity thereafter, 2D:4D has been suggested as an indirect marker for the organizational (permanent) masculinizing effects of prenatal androgens on the brain, behavior, and physique (Manning, 2002).

Given this hypothesis is correct, it would straightforwardly allow for comfortable (non-invasive) tests of the biological bases of behavior, cognition, personality, and other individual difference variables that show sex differences and therefore conceivably are partly influenced by prenatal sex-hormone action. Unsurprisingly, this idea has attracted wide interest, resulting in a sizeable literature now approaching 250 reports (partial reviews: Manning, 2002; Putz, Gaulin, Sporter, & McBurney, 2004).

2D:4D research also has a number of problems, including usually modest effects, preponderantly unreplicated findings, statistically underpowered (and thus type I error-prone) studies reporting positive findings, a spate of recently published nonreplications, and single-sample studies (without replication attempts) still being the norm.

Fecha recepción: 1-10-08 • Fecha aceptación: 11-1-09
Correspondencia: Martin Voracek
Dept. of Basic Psychological Research - School of Psychology
University of Vienna - Liebiggasse, 5, Rm 03-46
A-1010 Vienna (Austria)
E-mail: martin.voracek@univie.ac.at

Building on these combined facts, concerns of possible publication bias in 2D:4D research have been expressed (Putz et al., 2004).

This possibility seems not remote. Publication bias was clearly demonstrated for fluctuating asymmetry research (Palmer, 2000), a field sharing several features with 2D:4D research (e.g., small effects, measurement and methodology issues, underpowered studies, replication failures, rapid growth of the literature, a few productive labs, and competitive publishing).

To give one example, an initial report (Csathó, Osváth, Bicsák, Karádi, Manning, & Kállai, 2003), based on merely 46 female Hungarian students, suggested associations of 2D:4D with sex-role orientation. According to ISI Web of Knowledge, this report ranks among the top 15 of most frequently cited papers of 2D:4D research. However, five subsequent reports, including large-scale and multi-sample evidence (Hampson, Ellis, & Tenk, 2008; Lippa, 2006; Rammsayer & Troche, 2007; Schmukle, Liesenfeld, Back, & Egloff, 2007; Troche, Weber, Hennigs, Andresen, & Rammsayer, 2007), and at least six unpublished theses (Evardone, 2006; Fawcett, 2003; Meingaßner, 2003; Robinson, 2005; Schicker, 2005; Vonnahme, 2005) have failed to replicate such associations. Two further studies with positive findings (Beech & Mackintosh, 2005; Scarbrough & Johnston, 2005) were based on similarly small samples (*N*s= 60 and 41). On the whole, there is more evidence against associations of 2D:4D with sex-role orientation than support for these.

In this context, one study (Schmukle et al., 2007) is of particular importance, because it was hypothesized that implicit (automatic or nonconscious) measures, as opposed to explicit (controlled or conscious) measures, might entertain more reliable associations with 2D:4D. This is because explicit mental representations, as gauged by self-report measures, exploit propositional structures that are accessible through introspection, malleable, and thus reflect more recent events/experiences, whereas implicit mental representations build on associative structures that are unaccessible through introspection, less malleable, and thus may reflect earlier events/experiences (Fazio & Olson, 2003).

Using the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), a widely applied experimental method for measuring implicit social cognition (Lane, Banaji, Nosek, & Greenwald, 2007), Schmukle et al. (2007) indeed obtained stronger associations (significant for men, but not significant for women) of 2D:4D with sex-role orientation, when measured implicitly rather than explicitly. However, this evidence again was based on a small (39 men and 42 women) single-sample design without replication attempt.

There are several generic approaches for clarifying conflicting research literatures, most of which have already been tried within 2D:4D research, including meta-analyses (e.g., Puts, McDaniel, Jordan, & Breedlove, 2008), individual-data reanalyses of studies (e.g., McFadden, Loehlin, Breedlove, Lippa, Manning, & Rahman, 2005), and multi-sample approaches (e.g., Troche et al., 2007). Replicability of findings, a straightforward scientific criterion, remains the *via regis* for establishing and furthering knowledge (Hendrick, 1990). For clarity, replications are differentiated into exact (or direct) vs. systematic (or conceptual) ones. Exact replications imply repeating studies using identical procedures, thus supporting (if successful) or contesting (if unsuccessful) a theory. Systematic replications use theoretical, procedural, or operational changes to the original study, thus extending (if successful) or limiting (if unsuccessful) a theory.

In the present research, we applied both types of replications. The rationale was as follows: if candidate individual difference variables indeed show stronger or more consistent associations with 2D:4D when measured implicitly rather than explicitly (Schmukle et al., 2007), this should generalize to other traits than sex-role orientation. Several studies have investigated associations of 2D:4D with types of sexually differentiated aggression (Austin, Manning, McInroy, & Mathews, 2002; Bailey & Hurd, 2005; Benderlioglu & Nelson, 2004; Coyne, Manning, Ringer, & Bailey, 2007; Gallup, White, & Gallup, 2007; Hampson et al., 2008; Kuepper & Hennig, 2007; McIntyre, Barrett, McDermott, Johnson, Cowden, & Rosen, 2007; Millet & Dewitte, 2007). Of note, their findings appear less conflicting than the 2D:4D literature about sex-role orientation reviewed above and thus this line of inquiry possibly is more fruitful. A detailed review of these accounts is beyond the intended scope of the present work, but, suffice it to say, that the majority of the studies cited above yielded theory compliant within-sex associations of 2D:4D with various measures of aggression. Also, there is evidence for an early emergence of sex differences in aggression and related traits (e.g., childhood play) that are influenced by differences in early exposure to androgens (Hines, 2008). We therefore selected aggression (measured explicitly vs. implicitly) as the target trait for investigating associations with 2D:4D. This attempt of demonstrating generalizability constituted systematic replication.

As a further procedural extension, we considered two additional measures of hand anatomy that similarly are thought to be putative pointers to sex-hormonal exposure. First, D_{R-L}, the directional asymmetry (right-minus-left difference) in 2D:4D, typically is lower in men than in women (Manning, Churchill, & Peters, 2007; Voracek, Dressler, & Manning, 2007) and has therefore been proposed as another putative marker for prenatal androgen levels (Manning, 2002, pp. 21-22). For example, Coyne et al. (2007) found D_{R-L} to be negatively correlated with indirect aggression in women. Second, absolute finger length (2D and 4D), for which within-sex correlations with adult height are only moderately positive (merely hovering around .50; Lippa, 2003), seems to reflect pubertal-adolescent androgen levels, because the sex difference therein is negligible prior to puberty, but very large thereafter (Lippa, 2006). Because of the only loose association with adult height and further because of evidence for significant cross-talk of the androgen receptor (the target of testosterone) with growth-factor signalling pathways (Nieschlag, Behre, & Nieschlag, 2004), measures like absolute finger length cannot be regarded as pointers to growth-hormone exposure exclusively, but may well be appropriate markers for pubertal-adolescent testosterone exposure. Up to now, the measure of absolute finger length has only rarely been considered in 2D:4D research (e.g., Jackson, 2008).

We refrained from basing conclusions on findings yielded from a single sample, even when this was considerably larger than the one of the only predecessor study (Schmukle et al., 2007). We therefore collected data from a second, independent sample, using identical procedures. That way, we conducted an exact replication of our own research.

Methods

Participants

Sample I were 114 men and 130 women, ages 16 to 59 years (M= 27.5, SD= 10.3). Sample II were 101 men and 132 women, ages 15 to 75 years (*M*= 32.3, *SD*= 13.8). Participants were self-reported heterosexuals and native Austrians from the general population who volunteered for this research.

Instruments

Explicit aggression measures. The German form (Herzberg, 2003) of the most widely used self-report measure of aggression (Buss-Perry Aggression Questionnaire, BPAQ; Buss & Perry, 1992) was administered. This assesses the trait facets Physical aggression, Verbal aggression, Anger, and Hostility (8, 5, 6, and 8 items) on 5 point scales (-2: *extremely uncharacteristic of me*, +2: *extremely characteristic of me*). Following standard practice in IAT research (Greenwald et al., 1998), aggression was also explicitly assessed with a feeling thermometer (0 to 100 degrees).

Implicit aggression measure. IATs are now widely used keyboard-based rapid sorting tasks of stimuli presented on computer screens, whereby response latencies are analyzed to indirectly measure the relative strengths of automatic, nonconscious associations between paired concepts (commonly, between categories *Pleasant* vs. *Unpleasant*, coupled with two distinctive attributes). Sorting is easier (faster and less error-prone) when two concepts sharing a response key are strongly automatically associated than when less so (Greenwald et al., 1998). In the traditional IAT, genuinely intrapersonal associations are contaminated with extrapersonal (cultural) ones, because categories *Pleasant* vs. *Unpleasant* are likely perceived as normative. To reduce this contamination, personalized IAT variants (Olson & Fazio, 2004) substitute these with idiosyncratic categories (*I like* vs. *I don't like*) and use evaluation-laden stimuli with strong individual preferences, but no social consensus. A considerable body of evidence attests to the predictive validity of IAT measures using a variety of attitudes and self-concepts (for reviews, see Fazio & Olson, 2003; Greenwald & Nosek, 2001).

A personalized aggression IAT was assembled as follows: stimuli for the idiosyncratic categories (*I like* vs. *I don't like*: airplanes, beer, cleaning house, cigarettes, coffee, country music, darkness, disco, eating, football, garlic, jogging, Monday, motorcycles, opera, romance novels, spinach, storms, television, tequila) were adopted (Olson & Fazio, 2004) and those for the attributes (*Aggression*: anger, danger, fight, hothead, pain, provocation, punch, revenge, violence, weapon; *Peace*: calmness, freedom, health, heaven, joy, love, loyalty, pleasure, relaxation, tenderness) selected by the authors.

For a schematic overview, see Table 1. Stimuli from the categories (*I like* vs. *I don't like*) and the attributes (*Aggression* vs. *Peace*) were presented in 5 blocks. Practice blocks (1, 2, and 4) introduced concepts and stimuli (with each word presented twice). Blocks 1 and 4 were similar, except that attribute position was switched. Critically, concepts were differentially paired in Blocks 3 and 5, in order to measure their degree of association. Order of these blocks was not counterbalanced to maximize individual differences measurement and thus to enable unconstrained target-trait correlations (Egloff & Schmukle, 2002).

Table 1
IAT task sequence

Block	Trials	Block type	Concept label	
			Left key assignment	Right key assignment
1	40	Practice	*Peace*	*Aggression*
2	40	Practice	*I like*	*I don't like*
3	80	Test	*Peace + I like*	*Aggression + I don't like*
4	40	Practice	*Aggression*	*Peace*
5	80	Test	*Aggression + I like*	*Peace + I don't like*

The aggression IAT was run on laptops, with concept labels presented in the right and left upper screen corners and stimuli midway (intertrial interval: 150 ms). Participants were instructed to react as quickly as possible and used keys «E» vs. «I» for sorting to the left vs. right. Response latencies between stimulus presentation and keystroke were recorded throughout Blocks 1 to 5 in ms.

Finger-length measures. Following published 2D:4D ascertainment standards (Voracek, Manning, & Dressler, 2007), flatbed-scanned images of participants' right and left palms were produced. From high-resolution laser printouts of scanned images, length of 2D and 4D was measured triplicate, using the same image, from the middle of the flexion crease proximal-most to the palm to the fingertip by experienced, mutually blinded investigators (not the authors), using digital vernier calipers accurate to 0.01 mm. Average measurements were used to calculate right-hand and left-hand digit ratio (R2D:4D and L2D:4D).

Procedure

After reporting basic demographic information, participants completed the IAT, the BPAQ, and the feeling thermometer in quiet facilities. Presenting an IAT prior to corresponding explicit measures minimizes carry-over or priming effects (if any) among measures, because implicit measures by their nature are less likely to elicit such effects (Egloff & Schmukle, 2002). Finally, palm scans were taken and participants thanked and debriefed.

Data analysis

Calculation of individual IAT results followed standard practice, namely the originally proposed IAT effect algorithm (for full procedural details, see Greenwald et al., 1998) which continues to be used by a clear majority of IAT studies. Briefly, the IAT effect (in units of logged milliseconds) is the difference between the mean natural log transformed valid response latencies from Blocks 5 and 3, whereby higher (positive) values index less and lower (negative) values more aggression (i.e., compared with the explicit aggression measures, the IAT was reversely scored). Following Olson and Fazio (2004), participants with > 20% errors in attribute trials were discarded from analysis (see Table 2 note; by definition, response errors in idiosyncratic category trials are undetectable). Also following standard practice (Greenwald et al., 1998), calculation of the reliability of the IAT was based on the 78 response-time differences from the trials of Block 5 and 3 (the first two out of the 80 trials of these blocks are omitted).

Interobserver repeatabilities of finger-length measurements were assessed with single-score intraclass correlation coefficients (*ICC*; two-way mixed-effects model with absolute-agreement definition; Voracek et al., 2007). Significance was set to $p<.05$ (two-tailed), and p_{rep}, the probability of replicating an effect (Killeen, 2005), was calculated for selected findings for demonstration purposes.

Results

Sex differences in variables

Descriptive statistics and tests for sex differences are shown in Table 2. Replicating previous evidence (Manning, 2002) in both samples, men had lower 2D:4D than women. This sex effect was small-to-medium, somewhat weaker for L2D:4D than for R2D:4D, and nearly absent for D_{R-L}. Sex differences in absolute finger length were large, but invariably larger for 4D than for 2D, which pattern is the basis for the sex effect in 2D:4D.

Implicitly measured aggression was somewhat (but not significantly) stronger in men than in women in both samples. Explicitly measured aggression showed no sex effect (feeling thermometer) or generally small ones that were inconsistent across samples (BPAQ). Specifically, significantly higher Physical aggression scores of men relative to women in Sample I failed to replicate in Sample II, and significant sex differences in Anger scores were directionally inconsistent across samples.

Reliabilities of variables

These are displayed in Table 3 (diagonal entries). Consistent with prior evidence, in both samples finger-length and digit ratio measurements were highly repeatable (Voracek et al., 2007), internal consistency figures of the IAT high (Lane et al., 2007), and those of BPAQ subscales tallied to reference values (Herzberg, 2003).

Sex-specific associations among variables

All interrelations of variables are set out in Table 3 (off-diagonal entries). As expected, correlations among absolute finger lengths and digit ratios were strongly positive (upper-left corner entries) in both samples. Also, correlations among BPAQ

Table 2
Sex differences in variables under study

Variable	Men *M* (*SD*)	Women *M* (*SD*)	*t*	*p*	*d*
Digit ratio					
R2D:4D	0.957 (0.032)	0.969 (0.032)	-3.02	.003	-0.37
	0.958 (0.029)	0.968 (0.029)	-2.70	.007	-0.34
L2D:4D	0.956 (0.032)	0.967 (0.029)	-2.67	.008	-0.36
	0.956 (0.026)	0.964 (0.031)	-2.16	.03	-0.28
D_{R-L}	0.001 (0.019)	0.003 (0.026)	-0.65	.52	-0.09
	0.002 (0.027)	0.004 (0.031)	-0.51	.61	-0.07
Absolute finger length (mm)					
R2D	74.20 (3.99)	68.39 (4.43)	10.69	<.001	1.37
	72.43 (4.04)	67.55 (4.44)	8.64	<.001	1.14
R4D	77.59 (4.43)	70.61 (4.89)	11.66	<.001	1.49
	75.64 (3.96)	69.81 (4.71)	10.02	<.001	1.32
L2D	74.15 (4.33)	67.82 (4.52)	11.14	<.001	1.43
	72.59 (4.55)	67.09 (4.26)	9.48	<.001	1.25
L4D	77.59 (4.58)	70.20 (4.82)	12.24	<.001	1.57
	75.95 (4.73)	69.63 (4.86)	9.96	<.001	1.32
Average finger length	75.88 (4.01)	69.25 (4.49)	12.09	<.001	1.56
	74.15 (3.97)	68.52 (4.17)	10.44	<.001	1.38
Aggression measures					
IAT (log ms)	0.26 (0.22)	0.31 (0.24)	-1.57	.12	-0.22
	0.27 (0.24)	0.30 (0.22)	-1.04	.30	-0.13
BPAQ-Physical aggression	-0.98 (0.73)	-1.34 (0.57)	4.27	<.001	0.55
	-0.31 (0.51)	-0.29 (0.55)	-0.19	.85	-0.04
BPAQ-Verbal aggression	0.03 (0.64)	0.05 (0.60)	-0.22	.82	-0.03
	-0.37 (0.66)	-0.52 (0.63)	1.66	.10	0.23
BPAQ-Anger	-0.61 (0.71)	-0.40 (0.71)	-2.37	.02	-0.30
	-0.87 (0.56)	-1.09 (0.58)	2.90	.004	0.39
BPAQ-Hostility	-0.58 (0.70)	-0.49 (0.68)	-1.02	.31	-0.13
	-1.13 (0.51)	-1.29 (0.52)	2.41	.02	0.31
BPAQ-Total	-0.59 (0.50)	-0.62 (0.45)	0.46	.65	0.06
	-0.69 (0.43)	-0.81 (0.42)	2.13	.03	0.28
Feeling thermometer (degrees)	25.54 (20.71)	26.33 (18.99)	-0.31	.76	-0.04
	24.98 (21.48)	23.35 (18.50)	0.62	.54	0.08

First-line entries= Sample I (114 men, 130 women; for IAT: 95 and 107), second-line entries= Sample II (101 men, 132 women; for IAT: 79 and 120). *d*= Cohen's *d* (average group difference [male mean minus female mean] divided through square root of weighted mean of group variances). BPAQ= Buss-Perry Aggression Questionnaire (mean item responses). *P* values are two-tailed

Table 3
Sex-specific associations among variables under study

	R2D:4D	L2D:4D	D_{R-L}	R2D	R4D	L2D	L4D	Average	IAT	Physical	Verbal	Anger	Hostility	Total	Feeling
R2D:4D	**.92*****	.81***	.31**	.20*	-.39***	.15	-.32**	-.11	.04	-.04	.01	.10	.06	.04	.02
	.83***	.50***	.58***	.37***	-.18‡	.23*	.02	.12	.03	-.23*	-.01	-.05	.01	-.10	-.06
L2D:4D	.66***	**.89*****	-.30**	.20*	-.29**	.26**	-.31**	-.05	.14	.02	-.07	.09	.14	.08	-.03
	.49***	**.86*****	-.42***	.20*	-.07	.22*	-.21*	.04	.09	-.11	.05	-.15	-.04	-.08	.05
D_{R-L}	.50***	-.33***	**.75*****	.01	-.17‡	-.18‡	-.01	-.10	-.17‡	-.11	.14	.02	-.14	-.06	.09
	.45***	-.57***	**.71*****	.20*	-.12	.03	.21*	.09	-.07	-.13	-.06	.08	.05	-.02	-.11
R2D	.13	.13	.02	**.99*****	.82***	.86***	.73***	.92***	.03	-.08	.09	-.01	-.01	-.02	-.03
	.17*	-.01	.17*	**.98*****	.85***	.80***	.72***	.91***	.01	-.19‡	-.08	-.12	-.02	-.13	-.05
R4D	-.36***	-.20*	-.23**	.88***	**.99*****	.72***	.88***	.93***	.01	-.05	.08	-.06	-.03	-.02	-.03
	-.27**	-.22*	-.02	.90***	**.98*****	.72***	.75***	.89***	-.01	-.08	-.07	-.10	-.03	-.09	-.02
L2D	-.01	-.16‡	-.18*	.95***	.89***	**.99*****	.84***	.92***	.08	-.06	.01	.03	.11	.03	-.01
	.03	.06	-.03	.74***	.71***	**.99*****	.91***	.94***	-.05	-.11	-.03	-.10	-.03	-.09	-.02
L4D	-.29**	-.29**	-.04	.86***	.95***	.90***	**.99*****	.94***	-.01	-.07	.05	-.01	.03	-.01	.02
	-.20*	.41***	.23**	.70***	.75***	.89***	**.99*****	.93***	-.09	-.06	-.05	-.04	-.01	-.05	-.04
Average	-.15‡	-.06	-.11	.95***	.97***	.97***	.97***	**.99*****	.03	-.07	.06	-.01	.03	-.01	-.01
	-.08	.17*	.10	.91***	.92***	.91***	.91***	**.99*****	-.05	-.12	-.06	-.10	-.02	-.10	-.04
IAT	-.20*	-.10	-.13	-.08	.02	-.01	.04	-.01	**.95**	.03	-.10	-.05	.01	-.03	-.01
	-.15	-.04	-.10	-.09	-.02	-.03	-.01	-.04	**.96**	.04	-.06	-.43***	-.06	-.16	-.07
Physical	.05	.09	-.05	-.05	-.06	-.05	-.09	-.06	.05	**.77**	.20*	.46***	.27**	.74***	.48***
	.07	-.02	.08	-.06	-.09	.02	.03	-.03	.12	**.62**	.43***	.50***	.42***	.78***	.29**
Verbal	-.10	.04	-.17‡	-.05	.01	.01	-.01	-.01	-.07	.28**	**.50**	.34***	.20*	.51***	.40***
	.01	-.01	.01	-.03	-.04	.01	.01	-.02	-.06	.40***	**.53**	.44***	.43***	.72***	.28**
Anger	-.02	.09	-.13	-.08	-.07	-.05	-.09	-.07	.08	.42***	.41***	**.72**	.46***	.79***	.48***
	-.06	-.11	.06	-.17*	-.14	-.12	-.05	-.13	.16	.57***	.24**	**.58**	.56***	.80***	.27**
Hostility	.08	.08	-.01	-.07	-.10	-.07	-.10	-.09	.09	.30**	.17‡	.40***	**.72**	.73***	.30**
	.08	.06	.02	-.09	-.13	-.08	-.10	-.11	.16‡	.48***	.28**	.50***	**.63**	.79***	.31**
Total	.02	.11	-.10	-.09	-.09	-.06	-.11	-.09	.07	.72***	.56***	.78***	.74***	**.82**	.58***
	.04	-.02	.06	-.12	-.13	-.06	-.05	-.10	.14‡	.84***	.60***	.77***	.78***	**.82**	.38***
Feeling	.06	.19*	-.14	.04	.01	.06	-.02	.02	.08	.54***	.22*	.32***	.19*	.45***	–
	.13	.03	.10	-.05	-.11	-.05	-.06	-.07	-.06	.43***	.10	.40***	.37***	.44***	–

First-line entries= Sample I, second-line entries= Sample II. Diagonal entries= *ICC*s (finger-length measurements) or Cronbach's α (aggression measures). Off-diagonal entries= Pearson correlations, above-diagonal entries= men, below-diagonal entries= women. ‡ $p<.10$, * $p<.05$, ** $p<.01$, *** $p<.001$ (two-tailed)

subscales and between these and the feeling thermometer were consistently positive (lower-right corner entries). Implicitly measured aggression was largely dissociated from the explicit aggression measures, one noteworthy exception being a highly significant correlation between IAT effect and Anger scores emerging only for men in Sample II.

As for sex-specific associations of finger-length measures with implicit or explicit aggression measures (upper-right and lower-left corner entries), crucially, out of 8 (finger-length measures) × 7 (aggression measures) × 2 (sexes) × 2 (samples)= 224 computed correlations only 4 were significant. Not a single of these associations replicated across samples or would have survived multiple-tests correction, and 3 of them were directionally opposite to theory, with the exception of a negative correlation between R2D:4D and Physical aggression for men in Sample II (r= -.23, $p<.05$). Overall, the proportion of significant correlations (4 / 224= 1.8%) was even below the expected error rate of statistical tests (5%, if $p<.05$), when the null hypothesis (no effect) actually is true.

Rerunning the correlational analysis using Spearman's rank-order correlation coefficient also failed to yield any discernible pattern of consistent, replicable, or theory compliant associations of digit ratios with explicit or implicit aggression measures (details omitted), thus suggesting that the nil findings of the main analysis were robust and not influenced by outliers.

Discussion

This study was only the second one to examine associations of 2D:4D with a candidate trait measured both explicitly and implicitly. Samples I and II combined were almost six times larger than in the predecessor study (Schmukle et al., 2007). Based on a subsample of 39 men, the predecessor reported a significant correlation of r = .33 between R2D:4D and an implicit (IAT-based) measure of gender self-concept (i.e., masculinity-femininity). Assuming that this correlation magnitudinally was typical (i.e., that associations between 2D:4D and implicitly measured target traits typically would yield *r* figures of about .30), the a priori estimate of statistical power (Elashoff, 2000) to detect this theoretical *r* value of .30 would have merely been 45% with two-sided testing and a sample size of 39. This suggests that the finding of Schmukle et al. (2007) may well be a chance finding. However, given sample sizes of 79, 95, 107, or 120 (as were available for the tests for within-sex associations of digit ratios with the IAT in Samples I and II of the present study; see Table 1), the a priori power estimates to detect an expected *r* value of .30 with two-sided testing amount to 76%, 84%, 88%, or 91%, respectively. These values of statistical power are sufficiently high and thus suggest that the present two studies were not statistically underpowered and that their nil findings are unlikely to be due to statistical type II errors.

The current findings are clear-cut. Overall, there was no indication of theory compliant, reliable, or replicable within-sex associations of putative markers for prenatal (2D:4D, $D_{R\text{-}L}$) or pubertal-adolescent (absolute finger length) androgen action with either implicit or explicit measures of aggression. The lack of reliable 2D:4D associations with BPAQ scores is in line with nil findings of 2D:4D studies that also used the BPAQ (Austin et al., 2002; Coyne et al., 2007). Some recent studies (Millet & Dewitte, 2007, 2008, 2009; van den Bergh & Dewitte, 2006) point to the possibility that contextual cues may moderate 2D:4D effects, particularly cues eliciting or priming aggressive versus non-aggressive (altruistic and prosocial) behavior. By its design, this eventuality could not be addressed in the present research. However, it is worth mentioning that (similar to other strands of evidence from 2D:4D research) at present such context-dependent interaction effects for associations of 2D:4D with target traits have only been reported by one research group (and thus await independent replication) and furthermore have only been tested within a limited range of behavioral responses, namely largely for economic decision behavior (which calls for the need to demonstrate generalizability of such assumed interaction effects to a wider range of behavior). In our study, 2D:4D was also not reliably associated with aggression measured implicitly with a personalized IAT variant. In this regard, our systematic replication (or generalizability test) of the findings of Schmukle et al. (2007) failed: it is therefore uncertain (and remains to be tested anew) whether 2D:4D really is stronger or more reliably or anyway associated with sexually differentiated (and thus conceivably hormonally influenced) candidate individual difference variables, when these are assessed implicitly rather than explicitly. Further, explicit and implicit aggression measures themselves were largely dissociated, a finding not uncommon in IAT research (Lane et al., 2007).

Diverse knowns of previous research replicated in both samples, thus attesting to effect robustness, procedural validity, and data typicality. For instance, measurements of finger length and digit ratios were highly repeatable, internal consistency figures of the IAT were high (see Mierke & Klauer, 2003) and those of BPAQ subscales dovetailed to published normative data for its German form. Sex differences in digit ratios were significant, small-to-medium sized, somewhat more pronounced for R2D:4D than for L2D:4D, noticeably smaller and nonsignificant (but directionally theory compliant) for $D_{R\text{-}L}$. Those in absolute finger length were large, but invariably larger for 4D than for 2D, affirming the position that it is men's relatively longer 4D that accounts for the sex effect seen in 2D:4D. BPAQ subscales were throughout positively interrelated and likewise throughout positively correlated with a simple, single-item explicit aggression measure of high face validity (feeling thermometer). The novel personalized IAT variant for assessing aggression showed potential to yield sex effects in the predicted direction, although magnitudinally these were small.

In contrast, the failure of sex effects in the BPAQ to replicate reliably and directionally consistent across samples suggests psychometric shortcomings of this instrument's internal structure (see Herzberg, 2003), which also have become apparent in several adaptations of the BPAQ in other languages (Dutch version: Meesters, Muris, Bosma, Schouten, & Beuving, 1996; Japanese version: Nakano, 2001) as well as in an alternative German version of the BPAQ (von Collani & Werner, 2005). For instance, in one of their samples, von Collani and Werner (2005) found significant, but small, sex differences in Physical aggression scores (men higher: d= 0.31) and Anger scores (women higher: d= -0.25), whereas no significant sex differences in Verbal aggression or Hostility scores. The sex difference in Physical aggression scores replicated in another sample and was medium-sized (men higher: d= 0.64), but the sex difference in Anger scores did not replicate, and, on the other hand, there was a significant sex difference in Verbal aggression scores (contrary to expectation, men higher: d= 0.22) in this second sample, which had been absent in the first sample. Also, the sex difference in total BPAQ scores in the second sample was, although statistically significant, magnitudinally modest (men higher: d= 0.21).

More generally, the current nil findings add to a growing number of failures in 2D:4D research to replicate initially reported findings from small-*N*, single-sample designs in subsequent investigations providing large-scale or multi-sample evidence or both (e.g., Medland, Loehlin, & Martin, 2008). Presently, the majority of 2D:4D research findings is unreplicated. Time will show which of these will turn out unreplicable. In this regard, the combined approach of systematic replication attempts of prior evidence, along with exact replication attempts of one's own research, as pursued here, appears beneficial and of general interest for the 2D:4D research program. In tandem with large-scale, multi-sample studies and meta-analyses, this should help sorting out veridical findings from those due to chance, selective publishing, or related phenomena. To achieve this goal, making original nil findings and replication failures visible will be of utmost importance. By the same token, one conceivable implication of the current design is to adjust publication acceptance criteria in this field, such that a successful replication of 2D:4D associations with target traits must be presented.

A final proposal concerns the reporting of replication statistics. This should assist in gauging the veridicality of findings. For example, calculated p_{rep} values for the sex effect in digit ratios observed in our samples (Table 2) ranged from 0.94 to 0.98, indicating a high degree of trustworthiness. Indeed, the sex effect in 2D:4D undoubtedly constitutes the most robust and replicable 2D:4D research finding so far. However, p_{rep} figures for observed correlations between finger-length and aggression measures were much lower (many of them < .50), representing a high degree of uncertainty to see the same effect replicated in another research report. Or in one's own next sample, as we did.

References

Austin, E.J., Manning, J.T., McInroy, K., & Mathews, E. (2002). A preliminary investigation of the associations between personality, cognitive ability and digit ratio. *Personality and Individual Differences, 33,* 1115-1124.

Bailey, A.A., & Hurd, P.L. (2005). Finger length ratio (2D:4D) correlates with physical aggression in men but not in women. *Biological Psychology, 68,* 215-222.

Beech, J.R., & Mackintosh, I.C. (2005). Do differences in sex hormones affect handwriting style? Evidence from digit ratio and sex role identity as determinants of the sex of handwriting. *Personality and Individual Differences, 39,* 459-468.

Benderlioglu, Z., & Nelson, R.J. (2004). Digit length ratios predict reactive aggression in women, but not in men. *Hormones and Behavior, 46,* 558-564.

Buss, A.H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology, 63,* 452-459.

Coyne, S.M., Manning, J.T., Ringer, L., & Bailey, L. (2007). Directional asymmetry (right-left differences) in digit ratio (2D:4D) predict indirect aggression in women. *Personality and Individual Differences, 43,* 865-872.

Csathó, A., Osváth, A., Bicsák, E., Karádi, K., Manning, J.T., & Kállai, J. (2003). Sex role identity related to the ratio of second to fourth digit length in women. *Biological Psychology, 62,* 147-156.

Egloff, B., & Schmukle, S.C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology, 83,* 1441-1455.

Elashoff, J.D. (2000). *nQuery Advisor® Version 4.0* [computer software]. Cork, Ireland: Statistical Solutions Ltd.

Evardone, M. (2006). *Sex differences in anxiety: Testing a prenatal androgen hypothesis using behavioral and physiological markers.* Unpublished thesis, Texas A&M University.

Fawcett, C.C. (2003). *Individual difference factors and men's attraction to female body features.* Unpublished thesis, Brock University, Canada.

Fazio, R.H., & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology, 54*, 297-327.

Gallup, A.C., White, D.D., & Gallup, G.G., Jr. (2007). Handgrip strength predicts sexual behavior, body morphology, and aggression in male college students. *Evolution and Human Behavior, 28,* 423-429.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Greenwald, A.G., & Nosek, B.A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48*, 85-93.

Hampson, E., Ellis, C.L., & Tenk, C.M. (2008). On the relation between 2D:4D and sex-dimorphic personality traits. *Archives of Sexual Behavior, 37,* 133-144.

Hendrick, C. (1990). Replications, strict replications and conceptual replications: Are they important? In J.W. Neuliep (Ed.): *Handbook of replication research in the behavioral and social sciences* (pp. 41-50). Corte Madera, CA: Select Press.

Herzberg, P.Y. (2003). [Psychometric evaluation and validity of the German translation of the Aggression Questionnaire by Buss and Perry]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 24*, 311-323.

Hines, M. (2008). Early androgen influences on human neural and behavioural development. *Early Human Development, 84,* 805-807.

Jackson, C. (2008). Prediction of hemispheric asymmetry as measured by handedness from digit length and 2D:4D digit ratio. *Laterality, 13,* 34-50.

Killeen, P.R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science, 16,* 345-353.

Kuepper, Y., & Hennig, J. (2007). Behavioral aggression is associated with the 2D:4D ratio in men but not in women. *Journal of Individual Differences, 28,* 64-72.

Lane, K.A., Banaji, M.R., Nosek, B.A., & Greenwald, A.G. (2007). Understanding and using the Implicit Association Test: IV. What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.): *Implicit measures of attitudes* (pp. 59-102). New York: Guilford Press.

Lippa, R.A. (2003). Are 2D:4D finger-length ratios related to sexual orientation? Yes for men, no for women. *Journal of Personality and Social Psychology, 85,* 179-188.

Lippa, R.A. (2006). Finger lengths, 2D:4D ratios, and their relation to gender-related personality traits and the Big Five. *Biological Psychology, 71,* 116-121.

Manning, J.T. (2002). *Digit ratio: A pointer to fertility, behavior, and health.* New Brunswick, NJ: Rutgers University Press.

Manning, J.T., Churchill, A.J.G., & Peters, M. (2007). The effects of sex, ethnicity, and sexual orientation on self-measured digit ratio (2D:4D). *Archives of Sexual Behavior, 36,* 223-233.

Manning, J.T., Scutt, D., Wilson, J., & Lewis-Jones, D.I. (1998). The ratio of 2nd to 4th digit length: A predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Human Reproduction, 13,* 3000-3004.

McFadden, D., Loehlin, J.C., Breedlove, S.M., Lippa, R.A., Manning, J.T., & Rahman, Q. (2005). A reanalysis of five studies on sexual orientation and the relative length of the 2nd and 4th fingers (the 2D:4D ratio). *Archives of Sexual Behavior, 34,* 341-356.

McIntyre, M.H., Barrett, E.S., McDermott, R., Johnson, D.D.P., Cowden, J., & Rosen, S.P. (2007). Finger length ratio (2D:4D) and sex differences in aggression during a simulated war game. *Personality and Individual Differences, 42,* 755-764.

Medland, S.M., Loehlin, J.C., & Martin, N.G. (2008). No effects of prenatal hormone transfer on digit ratio in a large sample of same- and opposite-sex dizygotic twins. *Personality and Individual Differences, 44,* 1225-1234.

Meesters, C., Muris, P., Bosma, H., Schouten, E., & Beuving, S. (1996). Psychometric evaluation of the Dutch version of the Aggression Questionnaire. *Behaviour Research and Therapy, 34,* 839-843.

Meingaßner, E. (2003). [*Somatic markers for in utero androgen levels and spatial abilities*]. Unpublished thesis, University of Vienna.

Mierke, J., & Klauer, K.C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85*, 1180-1192.

Millet, K., & Dewitte, S. (2007). Digit ratio (2D:4D) moderates the impact of an aggressive music video on aggression. *Personality and Individual Differences, 43,* 289-294.

Millet, K., & Dewitte, S. (2008). A subordinate status position increases the present value of financial resources for low 2D:4D men. *American Journal of Human Biology, 20,* 110-115.

Millet, K., & Dewitte, S. (2009). The presence of aggression cues inverts the relation between digit ratio (2D:4D) and pro-social behaviour in a dictator game. *British Journal of Psychology, 100,* 151-162.

Nakano, K. (2001). Psychometric evaluation on the Japanese adaptation of the Aggression Questionnaire. *Behaviour Research and Therapy, 39,* 853-858.

Nieschlag, E., Behre, H.M., & Nieschlag, S. (Eds.). (2004). *Testosterone: Action, deficiency, substitution.* Cambridge, UK: Cambridge University Press.

Olson, M.A., & Fazio, R.H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653-667.

Palmer, A.R. (2000). Quasi-replication and the contract of error: Lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annual Review of Ecology and Systematics, 31,* 441-480.

Puts, D.A., McDaniel, M.A., Jordan, C.L., & Breedlove, S.M. (2008). Spatial ability and prenatal androgens: Meta-analyses of congenital adrenal hyperplasia and digit ratio (2D:4D) studies. *Archives of Sexual Behavior, 37,* 100-111.

Putz, D.A., Gaulin, S.J.C., Sporter, R.J., & McBurney, D.H. (2004). Sex hormones and finger length: What does 2D:4D indicate? *Evolution and Human Behavior, 25,* 182-199.

Rammsayer, T.H., & Troche, S.J. (2007). Sexual dimorphism in second-to-fourth digit ratio and its relation to gender-role orientation in males and females. *Personality and Individual Differences, 42,* 911-920.

Robinson, A.L. (2005). *Second to fourth digit ratio does not correlate with dominance or masculinity in either males or females.* Unpublished thesis, Appalachian State University.

Scarbrough, P.S., & Johnston, V.S. (2005). Individual differences in women's facial preferences as a function of digit ratio and mental rotation ability. *Evolution and Human Behavior, 26,* 509-526.

Schicker, K. (2005). [*Digit ratio (2D:4D), sex-role typicality, normative sex-role orientation, and sex-specific conflict behavior*]. Unpublished thesis, University of Vienna.

Schmukle, S.C., Liesenfeld, S., Back, M.D., & Egloff, B. (2007). Second to fourth digit ratios and the implicit gender self-concept. *Personality and Individual Differences, 43,* 1267-1277.

Troche, S.J., Weber, N., Hennigs, K., Andresen, C.-R., & Rammsayer, T.H. (2007). The relationship of digit ratio (2D:4D) and gender-role orientation in four national samples. *Journal of Individual Differences, 28,* 78-87.

van den Bergh, B., & Dewitte, S. (2006). Digit ratio (2D:4D) moderates the impact of sexual cues on men's decisions in ultimatum games. *Proceedings of the Royal Society of London Series B, Biological Sciences, 273,* 2091-2095.

von Collani, G., & Werner, R. (2005). Self-related and motivational constructs as determinants of aggression: An analysis and validation of a German version of the Buss-Perry Aggression Questionnaire. *Personality and Individual Differences, 38,* 1631-1643.

Vonnahme, P. (2005). *Facial preferences of Caucasian males as a function of 2D:4D digit ratio and mental rotation ability.* Unpublished dissertation, New Mexico State University.

Voracek, M., Dressler, S.G., & Manning, J.T. (2007). Evidence for assortative mating on digit ratio (2D:4D), a biomarker for prenatal androgen exposure. *Journal of Biosocial Science, 39,* 599-612.

Voracek, M., Manning, J.T., & Dressler, S.G. (2007). Repeatability and interobserver error of digit ratio (2D:4D) measurements made by experts. *American Journal of Human Biology, 19,* 142-146.