



Hormones and personality: Testosterone as a marker of individual differences

Jennifer Guinn Sellers^{a,*}, Matthias R. Mehl^c, Robert A. Josephs^b

^a Department of Psychology, Green Mountain College, One College Circle, Poultney, VT 05764, USA

^b Department of Psychology, University of Texas at Austin, 1 University Station, A8000, Austin, TX 78712, USA

^c Department of Psychology, University of Arizona, USA

Available online 4 May 2006

Abstract

Recently, testosterone (T) has been linked to behaviors that are conceptually related to dominance as a personality characteristic. Although evidence for this association is growing, the psychometric properties of T as an individual difference variable have been largely neglected. For T to be considered a biological marker of dispositional dominance it is critical that it demonstrates high test–retest reliability and good convergent and discriminant validity. Two studies tested the temporal stability of salivary T in humans and the relationship between T and traditional measures of personality. Across both studies, test–retest reliability for T was high and comparable to the short-term stability of questionnaire-based and implicitly assessed personality assessment instruments. In being modestly correlated with self-reported dominance, T showed some evidence of convergent validity. In being statistically independent from conceptually unrelated personality constructs (such as Emotional Stability and Openness to Experience) it showed good evidence of discriminant validity. The findings strengthen the psychometric foundation for using T as a hormonal marker of individual differences.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Testosterone; Personality; Individual differences; Stability; Dominance; Status

1. Introduction

Psychologists have long sought to reveal the biological basis of personality (Canli, in press; Eysenck, 1967). Although the number of studies aimed at understanding the genetic

* Corresponding author. Fax: +1 802 287 8099.

E-mail addresses: sellersJG@greenmtn.edu (J.G. Sellers), Josephs@psy.utexas.edu (R.A. Josephs).

underpinnings of personality is quickly growing (for a review see Clark & Watson, 1999), research that examines the utility of other biologically based individual differences on predicting behavior, especially in humans, and remains scarce. De facto, the vast majority of personality research continues to rely on questionnaire-based self-report methods to assess human dispositions. Well-known problems surrounding the validity of self-reports (e.g., Dunning, Heath, & Suls, 2004; Paulhus & Vazire, in press; Wilson & Dunn, 2004) have recently spurred a growing interest in going beyond self-reports in personality assessment. Researchers have begun to use methods such as laboratory-based (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; Funder, Furr, & Colvin, 2000) or naturalistic observation (Mehl, Gosling, & Pennebaker, in press; Mehl & Pennebaker, 2003), linguistic analyses (Pennebaker & King, 1999), informant reports (Vazire, in press), or implicit assessments (e.g., Asendorpf, Banse, & Mücke, 2002; Brunstein & Schmitt, 2004; Schultheiss & Pang, in press) to make inferences about aspects of people's personalities.

Measures of hormonal activity constitute another excellent, albeit empirically unexplored, candidate for this new development because (a) many hormones can now be assessed efficiently and non-invasively from saliva samples¹ and (b) they bypass several validity concerns that plague self-reports.² This paper seeks to provide empirical support for the idea that testosterone (T) can serve as a marker of individual differences by demonstrating its relative temporal stability as well as aspects of convergent and discriminant validity.

Recently, a number of studies have found that basal T predicts behaviors related to acquiring and maintaining dominance. This has led some researchers to speculate about a theoretical connection between T and dominance as a personality characteristic (e.g., Josephs, Sellers, Newman, & Mehta, in press; Josephs, Newman, Brown, & Beer, 2003; Newman, Sellers, & Josephs, 2005; van Honk, Tuiten, & Verbaten, 1999). Yet, the assertion that basal T can serve as a biological marker of individual differences has so far been more implicit than explicit. Furthermore, the assertion has not been built on solid psychometric grounds. Although a growing number of studies have reported behavioral manifestations of basal T (cf. Josephs et al., in press; Josephs et al., 2003), surprisingly little research has documented its temporal stability as well as its convergent and discriminant validity with established measures of personality.

A small but growing empirical literature suggests a link between basal T and dominance behavior. Single measurements of T have been shown to predict a variety of behaviors related to dominance both within and outside of the laboratory (e.g., prison, Dabbs, Carr, & Frady, 1995; Dabbs & Hargrove, 1997; and the laboratory, Grant & France, 2001; van Honk et al., 1999, 2001). High T has been linked to dominance displays among female roommates (Cashdan, 1995) as well as more extreme behaviors such as rule-breaking among prison inmates (Dabbs, Frady, Carr, & Bersch, 1987), and getting into fights involving direct confrontations (Dabbs et al., 1995) as means of demonstrating dominance in prison. Modest correlations have also been observed between

¹ There are two types of circulating T. "Free" T is not bound to sex-hormone binding globulin, and is free to bind to receptors. As such, free T is the psychologically "active" fraction of T. "Total" T is a measure of both bound and unbound T. Measures of T in saliva only capture free T (Granger, Schwartz, Booth, & Arentz, 1999).

² To be sure, salivary T assessments possess their own unique set of validity concerns. See Granger, Shirtcliff, Booth, Kivlighan, and Schwartz (2004) for issues to be addressed when collecting, storing, and analyzing salivary T samples.

basal T and a person's tendency to view situations as a way to assert power or dominance over others (Schultheiss, Campbell, & McClelland, 1999; Schultheiss, Dargel, & Rhode, 2003). Thus, it appears that individuals high in T are motivated to achieve or maintain high status, and this desire influences their behavior (Archer, 2006).

Dominance strivings are not limited to high T individuals. Dubbed the “*mismatch effect*,” Josephs and colleagues (Josephs et al., in press; see also Newman et al., 2005) have recently shown that just as high T individuals display behaviors intended to achieve high status, low T individuals appear motivated to maintain *low status*. When experimentally placed into a high status position, low T individuals showed evidence of discomfort. They demonstrated an increase in emotional arousal and heart rate, evidence of cognitive distraction, and an implicit attention to status. Importantly, basal T predicted these effects *only* when status was threatened. Apparently, when there is neither a threat to status, nor an opportunity to change status, T does not predict behavior (cf. Newman et al., 2005; see also Kemper, 1990).

The effects of T on behavior also appear to be distinct from traditional measures of dominance. Whereas paper-and-pencil measures of dominance ask people to rate their typical behaviors across a wide variety of settings, recent studies linking T to dominance behaviors only find a situation-contingent effect, that is when status is threatened. For instance, when high T participants were being outperformed on intelligence tests during a rigged competition, they demonstrated evidence of cognitive distraction and reported increased emotional arousal. Participants did not show these effects when they were completing the same measures of intelligence alone. Furthermore, when the interaction between T and status was pitted against the dominance subscale of Jackson's (1967) Personality Research Form and Pratto, Sidanius, Stallworth, and Malle (1994) Social Dominance Orientation, the interaction between T and status remained significant whereas neither of the two personality measures nor the interaction between either personality measure and status accounted for a significant portion of the variance. It was only the interaction between T and status that predicted behavior (Josephs et al., in press). Thus, whereas T and traditional measures of dominance are conceptually related, each predicts, and in some instances fails to predict, a distinct set of dominance behaviors.

The evidence thus far suggests that basal T functions as a biological measure of dominance, namely by distinguishing people who prefer and feel most comfortable in high versus low status positions. In support of this assertion, in his review of the literature, Archer (2006) concluded that “People with higher existing levels of testosterone are more likely to show higher scores on a variety of different assessments of dominance... (p. 334).” To a large extent, however, this assertion was based on *behaviors* related to dominance. The relationship between T and questionnaire-based measures of dominance remains largely speculative. Therefore, one of the goals of the current paper is to test the relationship between T and questionnaire-based measures of dominance. Furthermore, despite the predictive power of T, the basic question of its temporal stability remains largely unanswered.

The temporal stability of T—a basic psychometric requirement for any individual difference measure—has not only empirically been largely neglected, but also conceptually been explicitly questioned (e.g., Mazur & Booth, 1998; O' Carroll, 1998). Some critics have asserted that instead of predicting behavior, T is best understood as a dependent variable (e.g., Mazur & Booth, 1998), arguing that as status increases, so too will T levels. According to this reciprocal theory, T levels rise after a successful dominance confrontation which in turn signals to the individual that further actions can be taken to maintain or enhance

status. Therefore, as status increases, so does T, and as T increases so too does an individual's behavior towards acquiring status. In contrast, after losing a dominance contest, T levels should decline and the individual should likewise flee from the competition to avoid further insult or loss of status.

There is evidence to support the assertion that after a successful dominance contest, T levels rise accordingly. In both intellectual and physical contests (e.g., chess, tennis), T levels have been shown to be higher in winners than in losers (e.g., Booth, Shelley, Mazur, Tharp, & Kittok, 1989; Mazur, Booth, & Dabbs, 1992; Mazur & Lamb, 1980). Importantly, this rise is only transitory and lasts an average of one to two hours. Additionally, it appears necessary that participants view the victory as a personal and relevant accomplishment. For participants who view a win as unimportant or as the result of chance, T levels remain stable or show only the expected diurnal decline (for a review see Archer, 2006). Additionally, rises after a victory have been shown to occur only in participants with a strong, personal power motive (Schultheiss et al., 1999) or a strong motivation to win (Suay et al., 1999). Based on these findings, Archer (2006) concluded that T levels only rise in individuals with an "...assertive style—seeking to dominate others...as opposed to [individuals with] a more inhibited style..." (p. 329). It is possible that these individuals "seeking to dominate others" are actually those with higher basal T levels. As of yet, this hypothesis remains untested. Nonetheless, it is clear that future research is needed to determine the research utility of both basal and changing T levels as each represent distinct, yet interrelated research phenomena. In the current set of studies, we address the viability of basal T as a *predictor* of behavior and thus one of our main goals is to determine the stability of salivary T.

The only direct test of the stability of salivary T in adults was reported by Dabbs (1990a; but see also Granger et al., 2004 for the stability of salivary T in adolescence). In accord with other research (e.g., Granger et al., 1999), Dabbs (1990a) found that T concentrations in men and women³ show circadian fluctuations with highest levels in the morning, and monotonic decreases throughout the day. T has also been found to fluctuate seasonally with nadirs in the spring (Dabbs, 1990b). Across two consecutive days, the average correlation between individual measurements at 5 times of day was $r = .64$ (computed simultaneously for men and women which is somewhat misleading given the large mean difference between the sexes). Finally, the correlation between T measurements in males over a period of nine weeks of study was .59 (.71 after removing outliers; Dabbs, 1990a).

Dabbs's (1990a) investigation on the stability of T is often cited by researchers as evidence for the temporal stability of T. Surprisingly though, this finding has never been replicated. Furthermore, Dabbs' (1990a) investigation (as well as *all* other studies that track T-levels over time) has been based on morning assessments (e.g., Burger, Dudley, Cui, Dennerstein, & Hopper, 2000; Harman, Metter, Tobin, Pearson, & Blackman, 2001). Most data on behavioral manifestations of T, however, have been collected in the afternoon, when T tends to be more stable (e.g., Granger et al., 1999). To date, there is *no* empirical evidence that speaks to the stability of T in the afternoon hours between Noon and 4 pm. This lack of evidence seriously undermines the assertion that T can serve as a marker of individual differences.

³ Although serum concentration is much higher than saliva concentration, *changes* in T concentration in saliva are equivalent to changes in serum concentration. Saliva and free serum T concentrations are very strongly correlated ($r = .97$). (Vittek, L' Hommedieu, Gordon, Rappaport, & Southren, 1985).

Few studies have found correlations between T and self-report measures of dominance. Grant and France (2001), for example, found a correlation of .24 between T and self-reported dominance in women. A finding paralleled behaviorally by Cashdan, 1995, who found a positive relationship between T and dominance displays among female roommates. Similarly, Udry and Talbert (1988) found that T correlated with self-report measures of dominance in a pubertal sample of males and females. More investigations are clearly needed, however, as well as samples that include adult male participants. Taken together, a basic construct validation of T requires a more thorough analysis of aspects of its convergent and discriminant validity.

Two studies were designed to address two basic questions surrounding the utility of T as a marker of individual differences. First, what is the short-term stability of T in men and women? Second, to what extent does T show empirical overlap with traditional, conceptually related and unrelated, personality measures? Although the construct validity of T as a measure of dominance has been well documented by researchers linking dominant *behaviors* to T, a thorough construct validation also comprises relating T to conceptually related (i.e., dominance) and unrelated (e.g., Emotional Stability) self-report measures of personality.

The following two studies were designed to address these issues. In the first study, T was measured in the afternoon for five consecutive days, in both sexes. The primary purpose of the first study was to assess the short-term stability of T. The second study again assessed T for both sexes, this time in two assessments spaced two days apart. To assess aspects of convergent and discriminant validity, participants completed a battery of self-report measures of personality. Two laboratories were used for the T assays to avoid potential confounds due to the unique procedures of any particular laboratory or assay (Dabbs et al., 1995).

2. Study 1

2.1. Methods

2.1.1. Participants

Twenty introductory psychology students (9 females, 11 males; age range 18–22) participated in the study for course credit.

2.1.2. Procedure and saliva analysis

To minimize the effects of diurnal fluctuations in T levels, all participants were assessed between the hours of Noon and 4 pm. Participants came to the lab for 5 consecutive days beginning on a Monday. Saliva collection was consistent with the procedure described by Granger et al. (1999). Participants first rinsed their mouths with water, and then chewed a piece of Trident original flavor sugarless gum to facilitate saliva secretion. After 5 min, they expelled their gum and drooled into a 1.8 mL vial. The saliva samples were stored in a –20 °C freezer before being shipped overnight on dry ice to Salimetrics, LLC (State College, PA). There, the T samples were analyzed using enzyme-linked immunoassay (sensitivity of the kit < 1.5 pg/ml; average intra-assay CV 6.5%; average inter-assay CV 7.5%).

2.2. Results and discussion

Likely due to blood contamination, the day-one T-levels of one female participant could not be determined and were thus excluded from the analyses. Statistics are reported for

male and female participants separately. The average T levels across the 5 days were 50.4 pg/ml for women and 144.2 pg/ml for men.

Average measure intraclass correlations based on two-way random effect models (ICC[2,k]; Shrout & Fleiss, 1979) were calculated to determine the stability of T. The reliability of the daily measures was very high for both men ($r = .94$, $p < .01$) and women ($r = .81$, $p < .01$). The average inter-item correlation across the 5 days was $r = .78$ for men and $r = .48$ for women. This difference was not significant, $z = .916$, $p = .18$. Table 1 shows the day-to-day bivariate correlations.

The high intraclass correlations indicate an impressive degree of short-term stability. High (low) levels of T on day one tended to co-occur with high (low) levels of T on all four subsequent days. Further, the magnitude of the stability of T-levels was comparable for men and women.

Although these data are promising, the small sample size may raise concerns about the generality of the findings. It is possible that participants with characteristics that are systematically linked to T instability (e.g., regular exercise) were underrepresented in the sample (cf. Väänänen & Mäntysaari, 2002). Additionally, the T levels in our sample were atypically high. Although other researchers have reported comparable T levels using the same enzyme-immunoassay procedure (e.g., Gray, Kahlenberg, Barrett, Lipson, & Ellison, 2002), substantially lower levels are more typical (e.g., Schultheiss et al., 2005).

In Study 2, we sought to test the validity of our stability findings. To do this, we changed assay procedures, opting for radio-immunoassay as opposed to enzyme-immunoassay, and used a different laboratory to conduct the assay. These steps were taken to determine if the effect obtained in Study 1 would survive a change of method and measurement. We also used a substantially larger sample in the hopes of achieving greater generality.

Further, Study 2 was designed to address issues of convergent and discriminant validity. Based on prior research (e.g., Grant & France, 2001) we expected T to show a modest positive correlation with a self-report measure of dominance. We did not expect a high correlation because prior research suggests that T and self-reported dominance capture distinctly different, albeit conceptually related, aspects of dominance (Josephs et al., in press; cf. Schultheiss & Brunstein, 2001). To clarify the degree of overlap between T and other important personality measures, we also sought to determine the relationship between T and the Big Five personality dimensions (John & Srivastava, 1999), and between T and

Table 1
Daily bivariate correlations among testosterone levels for men and women across 5 days

	Day 2	Day 3	Day 4	Day 5
Men ($n = 11$)				
Day 1	.76**	.68*	.73*	.92**
Day 2		.75**	.85**	.72*
Day 3			.89**	.78**
Day 4				.78**
Women ($n = 8$)				
Day 1	.47	.69	.83*	.49
Day 2		.37	.51	.21
Day 3			.36	.21
Day 4				.68

* $p \leq .05$.

** $p \leq .01$.

Masculinity–Femininity (e.g., McHale, Shanahan, Updegraff, Crouter, & Booth, 2004). Because assertiveness is a subset of extraversion, and assertiveness is also conceptually related to dominance, an argument could be made to hypothesize that T should correlate modestly with extraversion. However, as T is more directly conceptually related to dominance, and past studies have only found modest correlations, we were not optimistic about finding a significant correlation between these two measures. Furthermore, given the lack of direct conceptual overlap between the remaining Big Five dimensions and T, we did not expect T to correlate with these measures of personality either.

3. Study 2

3.1. Methods

3.1.1. Participants

Seventy-two introductory psychology students (37 females, 35 males; age range 17–22) participated in the study for course credit.

3.1.2. Procedure and saliva analysis

Participants arrived to the lab between Noon and 4 pm on two separate occasions, 48 h apart. The procedure for saliva collection was identical to the one described in Study 1. The vials were shipped overnight on dry ice to the Yerkes Research Center Core Endocrine Laboratory (Atlanta, GA) where they were analyzed using a modified radio-immunoassay kit produced by Diagnostic Systems Laboratories (Webster, TX) at the Yerkes Research Center (sensitivity 2–500 pg/ml; average intra-assay CV 8.7%; average inter-assay CV 17.6%).

3.1.3. Measures

Before the first saliva assessment, participants completed the following three widely used personality instruments: The Need for Power subscale of the Personality Research Form (Jackson, 1967; 16 items using a true-false response format, e.g., “I have little interest in leading others,” “I would like to be an executive with power over others”; Cronbach’s $\alpha = .84$), the Big Five Inventory (John & Srivastava, 1999, 44 items using a 5-point unipolar scale; e.g., “I see myself as someone who is talkative,” “... full of energy” for Extraversion, “I see myself as someone who is generally trusting,” “... has a forgiving nature” for Agreeableness, “I see myself as someone who worries a lot,” “... can be moody” for Emotional Stability; average $\alpha = .89$), and the Masculinity (8 items using a 5-point bipolar scale; e.g., “not at all independent” vs. “very independent,” “not at all competitive” vs. “very competitive”; $\alpha = .73$) and the Femininity subscale (8 items using a 5-point bipolar scale; e.g., “not at all emotional” vs. “very emotional,” “not at all kind” vs. “very kind”; $\alpha = .75$) of the Personal Attributes Questionnaire (Spence & Helmreich, 1978).

3.2. Results

Three participants were dropped from the analyses because one or more of their T levels deviated more than 4 standard deviations from the mean, leaving 36 women and 33 men in the sample. The average T levels across the two measurements were 16.63 pg/ml for women and 91.31 pg/ml for men.

3.2.1. Stability

Simple test–retest correlations were computed to determine the stability in T levels over 48 h. Very high correlations emerged for both men, $r(33) = .69$, $p < .01$, and women, $r(36) = .72$, $p < .01$. The combined correlation, after standardizing T levels within sex, was $r(69) = .70$, $p < .01$. See Fig. 1.

3.2.2. Convergent and discriminant validity with self-report measures

The convergent and discriminant validity of T was assessed via simple bivariate correlations between participants' mean T level and their scores on the personality instruments. As predicted, T was reliably but modestly correlated with self-reported dominance, $r = .248$, $p = .04$. No other significant correlations emerged with the exception of conscientiousness in females ($r = -.40$, $p < .05$). T levels for male and female participants were standardized separately in order to control for sex differences in mean levels (see Table 2). Correlations were collapsed across sex unless there was a significant difference between the sexes.

4. Discussion

Two studies were designed to strengthen the psychometric foundation for using T as a marker of individual differences. The test–retest reliability of T was high over the course of 5 days (Study 1) as well as over a period of 48 h (Study 2). Further, the degree of stability

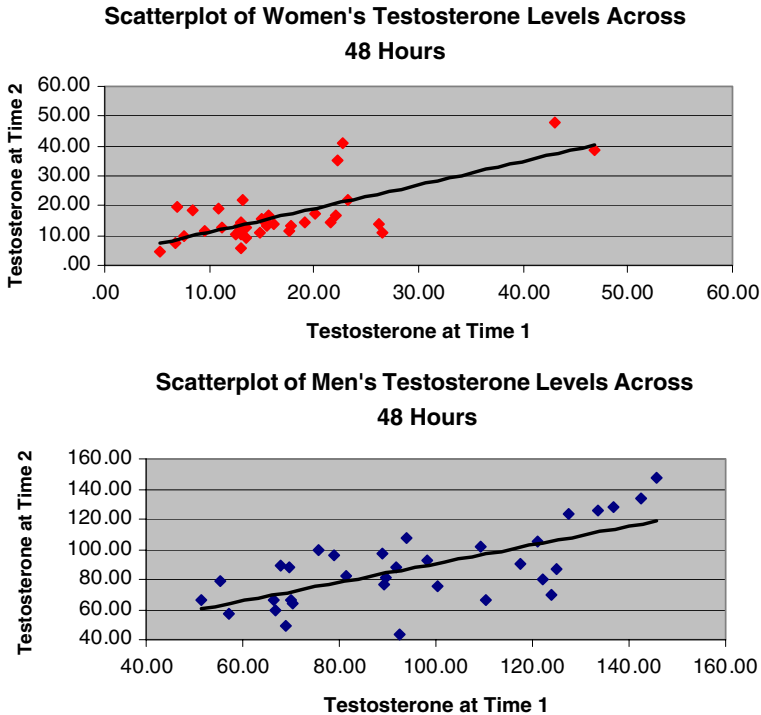


Fig. 1. Scatterplots of participants' testosterone levels (pg/ml) across 48 h.

Table 2

Correlation between testosterone and traditional personality measures

	Mean testosterone level across 48 h
<i>Personality Research Form</i>	
Dominance	.248*
<i>Big Five Inventory</i>	
Extraversion	.093
Agreeableness	-.024
Conscientiousness	
Men	.129 ($N = 33$)
Women	-.400* ($N = 36$)
Emotional Stability	.016
Openness to Experience	.083
<i>Personal Attributes Questionnaire</i>	
Masculinity	.071
Femininity	-.012

Note. Testosterone levels were standardized for men and women separately to control for sex differences in mean levels of testosterone. Correlations are collapsed across sex except for conscientiousness; * $p \leq .05$, $N = 69$ unless otherwise indicated.

was comparable for males and females. The robustness of the stability findings is further corroborated by the use of different testing laboratories (Dabbs, 1995) and different assay procedures (enzyme-immunoassay in Study 1, and radio-immunoassay in Study 2).

Combined with recent evidence that basal T predicts status-seeking behavior (Archer, 2006; Cashdan, 1995; Dabbs et al., 1987, 1995; Josephs et al., 2003; Josephs et al., in press; Newman et al., 2005; van Honk et al., 1999, 2001), this demonstration of the short-term stability of T is an important building block in construing T as a biological marker of individual differences in strivings for dominance (e.g., Newman et al., 2005; Schultheiss et al., 2005). Past research has shown that T is associated with implicit attention to status (Josephs et al., in press; Schultheiss et al., 1999; van Honk et al., 1999), dominant behaviors (Cashdan, 1995; Dabbs et al., 1987, 1995), and arousal to status threatening stimuli (Josephs et al., in press; Van Honk et al., 2001). The current study adds to these findings by demonstrating that these T-behavior links are likely not transitory, but rather to a large extent a reflection of stable individual differences in T. Future research needs to corroborate the impressive degree of short term stability documented in these studies by investigations into T's long-term stability over periods of months and years (but see Burger et al., 2000; Harman et al., 2001 for the effects of aging on T). In particular it is important to test the long-term stability of T against the magnitude of seasonal effects and other cyclic trends that are related to systematic T-relevant environmental changes.

How does the stability of T compare to the test-retest reliabilities obtained for other established personality measures? The Big Five Inventory (BFI; John & Srivastava, 1999) is one of the most widely used personality measures. A test-retest reliability of $r = .80$ over a two-week period has been reported for the BFI (Gosling, Rentfrow, & Swann, 2003). This is comparable to the stability of T found in the two studies here (.69 to .94). The identified stability of T also fares comparably with an implicitly assessed personality measure, the picture story exercise (or PSE; cf. McClelland, Koestner, & Weinberger, 1989). In the PSE, participants describe ambiguous pictures which trained coders then score for evidence of power, achievement and affiliation/intimacy imagery. In a recent meta-analysis,

Schultheiss and Pang, *in press* estimated that the average reliability for the PSE is .71 over the course of 1 day and .60 over the course of 1 week. It seems, then, that T demonstrates a similar degree of short term stability as other established explicit and implicit individual difference measures.

In Study 2, T demonstrated the predicted modest degree of convergence with a widely used self-report measure of dominance, the dominance scale of the Personality Research Form (Jackson, 1967). This finding is consistent with the correlation between T and dominance in women reported by Grant and France (2001). Furthermore, it mirrors behavioral data that suggest that T is a biological correlate of dominance (Archer, 2006; Josephs et al., 2003, *in press*; Newman et al., 2005). Finally, this finding is consistent with research on implicit motives which has documented a modest correlation ($r = .20$) between basal T and the power motive as measured by the PSE (Schultheiss et al., 1999). Future research is needed to identify the conceptual and empirical commonalities between these 3 distinct measurement approaches to assessing dominance and to test under which conditions each one will demonstrate unique predictive validity (cf. Roberts, Harms, Smith, Wood, & Webb, 2005).

Further, as predicted, T showed no reliable associations with conceptually unrelated personality measures such as the BFI and the PAQ Masculinity and Femininity scale (discriminant validity). Interestingly, T correlated negatively with conscientiousness among female participants. This finding was unpredicted and warrants further exploration. Also, T failed to show any positive relationship to self-reported Extraversion, the Big Five dimension that absorbs individual differences in assertiveness.

Undoubtedly, T is not as easy to measure as self-reported dominance. There is not only the added monetary expense, but also the significant time required to conduct the hormonal assay. Nonetheless, if we are to more fully understand the multi-faceted nature of personality, we need to broaden the conceptualization of personality traits. By broadening their operationalization of dominance from self-report to T, Josephs and colleagues showed that T has superior predictive validity over self-report measures of dominance in status relevant situations (Josephs et al., *in press*). Indeed, had Josephs et al. only relied on self-reported dominance, they would not have found the decline in cognitive performance or increase in arousal when dominance needs (as measured by T in their studies) were frustrated. They argued that T influences behavior only in certain situations. Thus, it (and other hormones) can be thought of as situationally-contingent personality measures. Self-reports, on the other hand, are conceptualized to predict behaviors regardless of the situation (e.g., if someone claims to be agreeable, he or she is expected to be agreeable across many if not most situations). Thus, the extra effort associated with measuring T and other hormonal markers of personality are justified on both conceptual grounds (hormones measures aspects of personality not tapped into by self-report measures) and on methodological grounds (in those situations in which hormonal measurements are appropriate, they have the potential to predict and explain behavior above and beyond self-reports).

Another assessment approach not investigated in the current set of studies is the implicit assessment of dominance motives. Consistent with the findings reported above, the implicitly assessed power motive does not map well onto self-reported dominance (Schultheiss & Brunstein, 2001; Schultheiss & Pang, *in press*), although the two constructs are undoubtedly conceptually related. Future investigations into these three distinct yet interconnected aspects of human dominance striving has the potential to yield a more thorough understanding of personality—the biological underpinnings (hormones), unconscious drives

(implicit motives), and people's understanding of who they are (personality self-concept). Adopting an individual difference perspective on T can only serve to enhance our understanding of this evolving field of research.

Taken together, these findings strengthen the psychometric basis for T as a marker of individual differences. They add empirically as well as conceptually to existing research on the interplay between hormones, personality, and social behavior. What is exciting about these data is the fact that hormonal markers have the potential to contribute as biological measures to the multi-faceted phenomenon of human personality (Canli, *in press*).

Acknowledgments

This research was supported by NSF Grants No. NSF199601580-002 and BCSO423405 to RAJ. We are grateful to Valerie Grant, Jamie Pennebaker and Matthew Newman for their helpful suggestions.

References

- Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience and Biobehavioral Reviews*, *30*, 319–345.
- Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. *Journal of Personality and Social Psychology*, *83*, 380–393.
- Booth, A., Shelley, G., Mazur, A., Tharp, G., & Kittok, R. (1989). Testosterone, winning and losing in human competition. *Hormones and Behavior*, *23*, 556–571.
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*, 599–614.
- Brunstein, J. C., & Schmitt, C. H. (2004). Assessing individual differences in achievement motivation with the Implicit Association Test. *Journal of Research in Personality*, *38*, 536–555.
- Burger, H. G., Dudley, E. C., Cui, J., Dennerstein, L., & Hopper, J. L. (2000). A prospective longitudinal study of serum testosterone, dehydroepiandrosterone sulfate, and sex hormone-binding globulin levels through the menopause transition. *Journal of Clinical Endocrinology and Metabolism*, *85*, 2828–2832.
- Canli, T. (Ed.). The biological bases of personality and individual differences. New York: Guilford, *in press*.
- Cashdan, E. (1995). Hormones, sex, and status in women. *Hormones and Behavior*, *29*, 354–366.
- Clark, L. A., & Watson, D. (1999). Temperament: a new paradigm for trait psychology. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 251–276). New York: The Guilford Press.
- Dabbs, J. M., Frady, R. L., Carr, T. S., & Besch, N. F. (1987). Saliva testosterone and criminal violence in young adult prison inmates. *Psychosomatic Medicine*, *49*, 174–182.
- Dabbs, J. M., Jr. (1990a). Salivary testosterone measurements: reliability across hours, days, and weeks. *Physiology & Behavior*, *48*, 83–86.
- Dabbs, J. M., Jr. (1990b). Age and seasonal variation in serum testosterone concentrations among men. *Chronobiology International*, *7*, 245–249.
- Dabbs, J. M., Jr. (1995). Reliability of salivary testosterone measurements: a multicenter evaluation. *Clinical Chemistry*, *41*, 1581–1584.
- Dabbs, J. M., Jr., Carr, T. S., & Frady, R. L. (1995). Testosterone, crime, and misbehavior among 692 male prison inmates. *Personality & Individual Differences*, *18*, 627–633.
- Dabbs, J. M., Jr., & Hargrove, M. F. (1997). Age, testosterone, and behavior among female prison inmates. *Psychosomatic Medicine*, *59*, 477–480.
- Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment: implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *69*, 106.
- Eysenck, H. J. (1967). *The biological basis of personality*. Baltimore, MD: University Park Press.
- Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The riverside behavioral Q-sort: a tool for the description of social behavior. *Journal of Personality*, *63*, 451–489.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the big five personality domains. *Journal of Research in Personality*, *37*, 504–528.

- Granger, D., Schwartz, E., Booth, A., & Arentz (1999). Salivary testosterone determination in studies of child health and development. *Hormones and Behavior*, *35*, 18–27.
- Granger, D. A., Shirtcliff, E. A., Booth, A., Kivlighan, K. T., & Schwartz, E. B. (2004). The “trouble” with salivary testosterone. *Psychoneuroendocrinology*, *29*, 1229–1240.
- Grant, V. J., & France, J. T. (2001). Dominance and testosterone in women. *Biological Psychology*, *58*, 41–47.
- Gray, P. B., Kahlenberg, S. M., Barrett, E. S., Lipson, S. F., & Ellison, P. T. (2002). Marriage and fatherhood are associated with lower testosterone in males. *Evolution and Human Behavior*, *23*, 193–201.
- Harman, S. M., Metter, E. J., Tobin, J. D., Pearson, J., & Blackman, M. (2001). Longitudinal effects of aging on serum total and free testosterone levels in healthy men. *Journal of Clinical Endocrinology and Metabolism*, *86*, 724–731.
- Jackson, D. N. (1967). *Personality research form manual*. New York: Research Psychologists Press.
- John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: history, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford Press.
- Josephs, R. A., Newman, M. L., Brown, R. P., & Beer, J. M. (2003). Status, testosterone, and human intellectual performance: stereotype threat as status concern. *Psychological Science*, *14*, 158–163.
- Josephs, R. A., Sellers, J. G., Newman, M. L., & Mehta, P. H. (in press). The mismatch effect: when testosterone and status are at odds. *Journal of Personality and Social Psychology*.
- Kemper, T. D. (1990). *Social structure and testosterone: Explorations of the socio-bio-social chain*. New Brunswick, NJ: Rutgers University Press.
- Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral Brain Sciences*, *21*, 353–397.
- Mazur, A., Booth, A., & Dabbs, J. M., Jr. (1992). Testosterone and chess competition. *Social Psychology Quarterly*, *55*, 70–77.
- Mazur, A., & Lamb, T. (1980). Testosterone, status, and mood in human males. *Hormones and Behavior*, *14*, 236–246.
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (in press). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*.
- Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: a psychometric analysis of students’ daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*, 857–870.
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, *96*, 690–702.
- McHale, S. M., Shanahan, L., Updegraff, K. A., Crouter, A. C., & Booth, A. (2004). Developmental and individual differences in girls’ sex-typed activities in middle childhood and adolescence. *Child Development*, *75*, 1575–1593.
- Newman, M. L., Sellers, J. G., & Josephs, R. A. (2005). Testosterone, cognition and social status. *Hormones and Behavior*, *47*, 205–211.
- O’Carroll, R. E. (1998). Placebo-controlled manipulations of testosterone levels and dominance. *Behavioral Brain Sciences*, *21*, 382–383.
- Paulhus, D. L., & Vazire, S. (in press). Self-report methods. In R. W. Robins, R. C. Fraley, R. Krueger (Eds.), *Handbook of research methods in personality psychology*. New York: Guilford Press.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of Personality and Social Psychology*, *77*, 1296–1312.
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: a personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*, 741–763.
- Roberts, B. W., Harms, P. D., Smith, J., Wood, D., & Webb, M. (2005). Methods in personality psychology. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association.
- Schultheiss, O. C., Campbell, K. L., & McClelland, D. C. (1999). Implicit power motivation moderates men’s testosterone responses to imagined and real dominance successes. *Hormones and Behavior*, *36*, 234–241.
- Schultheiss, O. C., & Brunstein, J. C. (2001). Assessment of implicit motives with a research version of the TAT: picture profiles, gender differences, and relations to other personality measures. *Journal of Personality Assessment*, *71*(1), 71–86.
- Schultheiss, O. C., Dargel, A., & Rhode, W. (2003). Implicit motives and gonadal steroid hormones: effects of menstrual cycle phase, oral contraceptive use, and relationship status. *Hormones and Behavior*, *43*, 293–301.
- Schultheiss, O. C., & Pang, J. S. (in press). Measuring implicit motives. In R. W. Robins, R. C. Fraley, R. Krueger (Eds.), *Handbook of research methods in personality psychology*. New York: Guilford.

- Schultheiss, O. C., Wirth, M. M., Torges, C. M., Pang, J. S., Villacorta, M. A., & Welsh, K. M. (2005). Effects of implicit power motivation on men's and women's implicit learning and testosterone changes after social victory or defeat. *Journal of Personality and Social Psychology*, *88*, 174–188.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing reliability. *Psychological Bulletin*, *86*, 420–428.
- Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents*. Austin, TX: University of Texas Press.
- Suay, F., Salvador, A., González-Bono, E., Sanchis, C., Martinez, M., Martinez-Sanchis, S., et al. (1999). Effects of competition and its outcome on serum testosterone, cortisol and prolactin. *Psychoneuroendocrinology*, *24*, 551–566.
- Udry, J. R., & Talbert, L. M. (1988). Sex hormone effects on personality at puberty. *Journal of Personality and Social Psychology*, *54*, 291–295.
- Väänänen, I., & Mäntysaari, M. (2002). Hormonal responses to daily strenuous walking during 4 successive days. *European Journal of Applied Physiology*, *88*, 12–127.
- van Honk, J., Tuiten, A., & Verbaten, R. (1999). Correlations among salivary testosterone, mood, and selective attention to threat in humans. *Hormones and Behavior*, *36*, 17–24.
- Van Honk, J., Tuiten, A., Hermans, E., Putman, P., Koppeschaar, H., Thijssen, J., et al. (2001). A single administration of testosterone induces cardiac accelerative responses to angry faces in healthy young women. *Behavioral Neuroscience*, *115*, 238–242.
- Vazire, S. (in press). Informant reports: a cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, doi:10.1016/j.jrp.2005.03.003.
- Vittek, J., L' Hommedieu, D., Gordon, G., Rappaport, S., & Southren, A. (1985). Direct radioimmunoassay (RIA) of salivary testosterone: correlations with free and total serum testosterone. *Life Sciences*, *37*, 711–716.
- Wilson, T. D., & Dunn, E. (2004). Self-knowledge: its limits, value, and potential for improvement. *Annual Review of Psychology*, *55*, 493–518.