# 3

# Assessment of Salivary Hormones

Oliver C. Schultheiss
Steven J. Stanton

## A Primer on Concepts and Measurement Issues in Behavioral Endocrinology

"Hormones" are messenger molecules that are released by specialized neurons in the brain and by glands into the bloodstream, and that carry a signal at the speed of blood to other parts of the body. Which specific responses they trigger in target organs depends on the receptors involved and the functions of the organs. For instance, the peptide hormone arginine vasopressin (AVP) regulates water retention in the body when it binds to receptors in the kidneys, but enhances episodic memory when it binds to receptors in the brain (e.g., Beckwith, Petros, Bergloff, & Staebler, 1987). Thus one hormone can drive several different physiological and psychological functions through its effects on several target organs.

Generally, two broad classes of hormonal effects on physiology and behavior are differentiated. "Organizational effects" are lasting influences that hormones exert on the organism, thus changing its shape and functional properties in subtle (and sometimes not so subtle) ways. Orga-

nizational hormone effects often occur during development or at times of significant hormonal flux, such as puberty. For instance, the development of the female and male body morphology is largely under hormonal control during fetal development, and deviations from typical-gendered body morphology are frequently the result of deviations in hormone production, conversion, or receptor action. In contrast to organizational effects, "activational effects" are those that hormones exert temporarily, without effecting lasting changes in the organism. For instance, due to peaking estradiol levels around the time of ovulation, women become more sensitive to sexual stimuli, as indicated by an enhanced pupillary response. This effect vanishes again after ovulation, when estradiol levels decrease (Laeng & Falkenberg, 2007). Psychologists interested in the role of hormones in psychological functions and phenomena most frequently study such activational effects of hormones on the brain, although the much more difficult documentation of organizational hormone effects on behavior is also gaining traction (e.g., Baron-Cohen, Lutchmaya, & Knickmeyer, 2006).

The relationship between hormones and behavior is *bidirectional*. Hormones can have a facilitating effect on behavior, such as when high levels of testosterone facilitate learning a behavior that elicits an angry face (Wirth & Schultheiss, 2007), increase cardiac responses to angry faces (van Honk et al., 2001), or increase aggressive responses on a point subtraction game (Pope, Kouri, & Hudson, 2000). Such hormone → behavior effects can be most conclusively demonstrated through experimental manipulation of hormone levels—a method that is covered in detail by van Honk (Chapter 4, this volume). Conversely, the situational outcome of a person's behavior, as well as the stimuli and events impinging on the person, can influence current hormone levels. This is the case when winning or losing a contest raises or lowers testosterone levels in male mammals, including humans (e.g., Gladue, Boechler, & McCaul, 1989; Mazur, 1985; Oyegbile & Marler, 2005); when encounters with an attractive member of the other sex have an impact on an individual's sex hormones (e.g., Graham & Desjardins, 1980; Roney, Lukaszewski, & Simmons, 2007); or when watching romantic movies leads to an increase in viewers' progesterone levels (Schultheiss, Wirth, & Stanton, 2004).

Because hormones have far-reaching and broad effects on physiology and behavior, their release is tightly controlled and monitored, primarily through negative feedback loops. For instance, circulating levels of the steroid hormone cortisol are monitored by the brain. If levels fall below a critical threshold, the hypothalamus releases corticotropin-releasing hormone (CRH), which in turn triggers the release of adrenocorticotropic hormone (ACTH). ACTH travels from the brain to the

cortex of the adrenals (small glands that sit on top of the kidneys), where it stimulates the release of cortisol. If rising levels of cortisol exceed a certain threshold, CRH release—and thus the subsequent release of ACTH and cortisol—are curtailed until cortisol levels fall below the critical threshold again, due to metabolic clearance. As a consequence of this negative-feedback-loop mechanism, many hormones are released in repeated bursts occurring every 30–120 min. Notably, hormones can also influence the release of different hormones. The quick (i.e., within minutes) testosterone increase in response to dominance challenges sometimes observed in men and other male primates (Mazur, 1985) is a good example. These rapid changes are the result of the stimulating effects of epinephrine and norepinephrine (NE), which are released within seconds after the onset of a situational challenge, on the testes (which produce testosterone in men). This effect is independent of the hypothalamic–pituitary–gonadal feedback mechanism normally involved in testosterone release (Sapolsky, 1987).

Besides acquiring a basic understanding of endocrine function, two issues are of particular concern to behavioral scientists who want to include endocrine measures in their research. First, current hormone levels are *multiply determined*, and in order to tease out the effects of interest (i.e., relationships between hormones and behavior), it is almost always necessary to control for, or hold constant, other influences on hormone levels. Chief among those influences is the strong circadian variation observed in many endocrine systems. Hormones like testosterone, estradiol, and cortisol start out at high levels in the morning and then decline through the course of the day. The variance generated by this effect can easily drown out whatever between-subjects differences one hopes to observe in an experiment if it is not taken into account—either by recording and adjusting for time of day, or by conducting all testing only at one time of day (e.g., in the afternoon). Effects of experimental factors on hormone levels are more likely to be observed in the afternoon than in the morning, presumably because hormone levels are still too close to their physiological ceiling and too variable earlier in the day (see Dickerson & Kemeny, 2004; Wirth, Welsh, & Schultheiss, 2006). On the other hand, individual differences in peak hormone levels measured in the morning may be a better predictor of behavioral responses to emotional stimuli than hormone measurements later in the day may be (Wirth & Schultheiss, 2007; see also Schultheiss et al., 2005).

Another important chronobiological influence on observed hormone levels is the menstrual cycle. The gonadal steroids estradiol and progesterone vary strongly as a function of cycle phase, and researchers therefore try to control hormone variance due to cycle stage—either by testing only women in a particular stage (e.g., the follicular phase), or

by asking female research participants to report the onset date of the last menses and average cycle length, in order to make a rough calculation later of what cycle phase each participant was in during the time of testing.

Besides chronobiological effects on hormone levels, drug status of research participants should also be taken into account (e.g., Daly et al., 2003; Hibel, Granger, Kivlighan, & Blair, 2006; Nadeem, Attenburrow, & Cowen, 2004). Probably the drug most frequently used by female participants is the oral contraception pill, which keeps endogenous gonadal steroid levels low throughout the menstrual cycle. Other, less frequently encountered drugs with profound effects on hormone levels include anabolic steroids, steroid-based anti-inflammatory medications, antidepressants, or drugs influencing the body's fluid levels. Appendix 3.1 is a screening questionnaire we frequently use in behavioral endocrinology studies to control for the most important extraneous influences on hormone levels.

The second issue of concern to a behavioral scientist who wants to use endocrine measures is how easy or difficult it is to assess a particular hormone. This in turn depends primarily on the biochemical properties of the hormone. *Peptide hormones* (i.e., short protein molecules composed of a small number of amino acids), such as insulin, AVP, ACTH, NE, and oxytocin (OXY), are large structures by molecular standards and therefore do not easily pass through cell membranes. As a consequence, they can only be measured in the medium or body compartment into which they have been released or actively transported. For instance, OXY released by the pituitary into the bloodstream can only be assessed in blood, but not in saliva (Horvat-Gordon, Granger, Schwartz, Nelson, & Kivlighan, 2005; but see Carter et al., 2007). Also, OXY concentrations measured in the body may not accurately reflect OXY levels in the brain, because they are released by different hypothalamic sites. Moreover, peptide hormones break down easily, and special precautions are necessary to stabilize their molecular structure after sampling. The other major class of hormones besides the peptides are *steroid hormones*, which are synthesized in the body from cholesterol. In contrast to peptide hormones, steroid hormones are highly stable, and in their free, bioactive form (i.e., not bound to larger proteins) they can pass through cell membranes, leading to roughly similar levels of the free fraction of a hormone across body compartments. This means that, for instance, cortisol levels measured in saliva are similar to (free) cortisol levels measured in blood or cortisol levels in the brain. For this reason, and because saliva sampling is much easier and less stressful for research participants than the collection of blood samples or spinal fluid samples (to get at hormone levels within the central nervous system [CNS]), salivary hor-

mone assessment has become the method of choice among behavioral endocrinologists and psychologists working with human populations (Dabbs, 1992; Hofman, 2001; Kirschbaum & Hellhammer, 1994).

One of the most frequently used methods for the assessment of salivary hormone levels is the radioimmunoassay (RIA). Although RIAs for many hormones can be purchased from commercial vendors, and a researcher therefore does not necessarily have to know how they are manufactured, we think it is useful to know what the term "radioimmunoassay" actually means. An "assay" is a procedure by which the concentration of an analyte (a hormone, in this case) is measured. "Immuno-" signifies that hormonal assays exploit the property of an organism's immune system to produce antibodies (i.e., proteins that adhere in a highly specific manner to alien compounds entering the body) in response to the injection of alien organic matter (antigens); in the case of a hormonal RIA, animals are injected with the human form of a given hormone, and the antibodies produced in response to the injected hormone are harvested from the animal's blood, purified, and used in the RIA to bind to the hormone content added to the assay. "Radio-" signifies that in RIAs, a fixed quantity of hormone molecules with radioactive labels (typically radioiodine [$^{125}$I]) is added to the assay, and these molecules compete with molecules from samples collected in participants for antibody-binding sites (see Figure 3.1 for a schematic overview of an RIA).

An RIA is therefore a special case of a competitive immunoassay, in which labeled and unlabeled hormones compete for binding sites until they settle into a binding equilibrium that depends only on the concentration of unlabeled hormone in the sample, since the amounts of labeled hormone and antibodies are fixed across the entire assay. Assays have also been developed that use enzymatic labels whose relative presence in a given sample is indicated by degrees of coloration, fluorescence, or luminescence. The chief advantage of such enzymatic immunoassays (EIAs) is that they do not require the use of radioactivity and thus the licensing of the assay facility, personnel training, and precise bookkeeping associated with the use of radioactive substances. The drawbacks of EIAs include complex assay protocols and relatively less accuracy and sensitivity than RIAs (see Raff, Homar, & Burns, 2002).

## Using Salivary Hormone Assays in Psychological Research: A Guided Tour

In this section, we illustrate the use of a hormonal assay in social neuroscience by describing, from start to finish, the procedures in a study

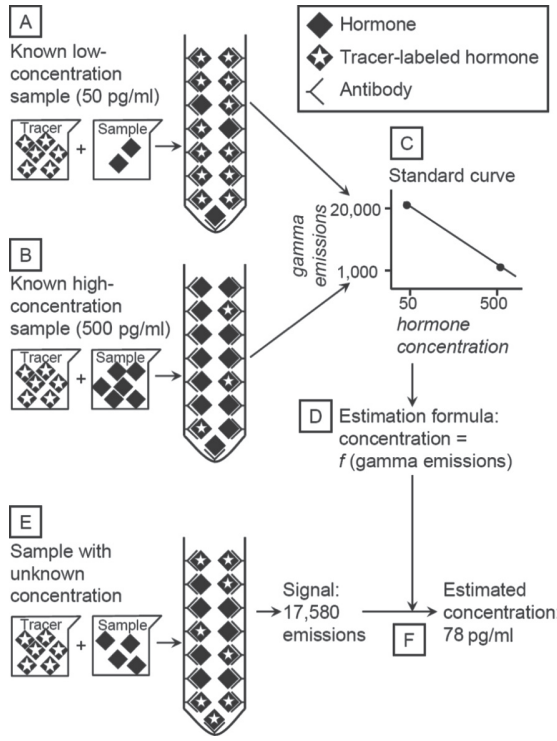



**FIGURE 3.1.** Schematic overview of a competitive radioimmunoassay (RIA). Each RIA contains, besides the samples whose hormone concentrations need to be determined, standards (or calibrators) with known amounts of hormone concentrations that cover the entire range of the hormone levels expected in the given medium (e.g., saliva) and population studied. In this simplified schematic, a standard with a low concentration (A) and a standard with a high concentration (B) are added to assay tubes containing antibodies that react with the hormone of interest. Radioactive tracer with a fixed concentration of radioisotope-labeled hormone is also added to the assay. Labeled and unlabeled hormones compete for antibody-binding sites until equilibrium is reached. Sample and tracer fluid is then discarded, leaving only antibody-bound hormone, both labeled and unlabeled, in the tube. Tubes are measured in a gamma counter, which provides a high gamma signal for the low-concentration sample (because tracer-labeled hormone is bound to the majority of antibodies) and a low gamma signal for the high-concentration sample (because more unlabeled than tracer-labeled hormone is bound to the antibodies). Interpolation between standards yields a standard curve (C) that allows estimating a given sample's concentration from the number of gamma emissions measured. The relationship between gamma emissions and sample concentration can be calculated and expressed through an estimation formula (D), based on a regression of concentration on gamma signal after appropriate linearization of both variables. This formula can then be applied to samples with unknown concentrations (E; e.g., participants' saliva samples), for which hormone concentrations can be estimated on the basis of the gamma signal (F).

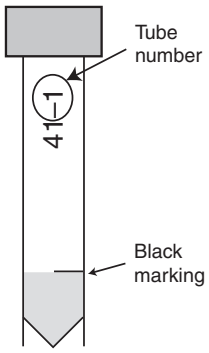
on the joint effects of implicit power motivation and winning or losing a dominance contest on testosterone changes in men (Schultheiss et al, 2005, Study 1). In the course of a 2-hour experiment, three saliva samples were collected before a dominance contest and three after the contest was over. Precontest samples were collected to obtain a baseline at the very start of the session (0 min); after the experimenter had announced that participants would compete against each other on a speed-based task (T2, 52 min); and after participants had imagined the ensuing contest from the winner's perspective through a guided imagery exercise (T3, 64 min, immediately before the contest). Samples at T2 and T3 were taken to examine the effects of verbal instructions versus experiential elaboration of these instructions, respectively (see Schultheiss & Brunstein, 1999; Schultheiss, Campbell, & McClelland, 1999). The T3 sample also served as the baseline covariate closest to the actual contest in later analyses. Postcontest samples were taken immediately after the contest (T4, 78 min) and then, at intervals of 15 min, twice more (T5 and T6) to explore the time course of power motivation × contest outcome effects. Sampling intervals of 15 min and longer were chosen because of the time it takes for steroid hormones to transfer from blood into saliva and be cleared out of it again; the relative sluggishness of the process does not allow for greater temporal resolution (see Riad-Fahmy, Read, Walker, Walker, & Griffiths, 1987). Both pre- and postcontest samples were collected while participants completed other tasks (e.g., mood questionnaires) on a personal computer (PC).

### *Sample Collection*

The goal of the saliva collection phase is to collect high-quality samples (i.e., samples free of contaminants) in a precisely identified sequence and with a sufficient amount of saliva to allow the measurement of all targeted hormones later on. To eliminate contaminants like blood or residues from a meal, we asked participants to refrain from eating and brushing their teeth for at least 1 hour before coming to the lab. After they had arrived, they were asked to rinse their mouths with some water over a sink. During the experimental session, the collection of each sample was coupled to the completion of a noncritical task (e.g., providing mood ratings on the PC). Precise instructions about which tube to use and how much saliva to collect were given by instruction screens on the PC that also featured illustrative pictures (see Figure 3.2). After participants had completed the task, the following instruction was presented on the screen: "Have you filled the tube to the marked line? If not, please continue to collect more saliva before moving on to the next task. If the tube is filled to the line, please put the lid on the tube, place it back on

the desk, and spit out the chewing gum. Press a key to move on to the next task." Similar instructions were given for each of the six saliva collections. Of course, these instructions can also be given orally by an experimenter if no computers are used. Whichever method is used, care should be taken that every participant fully understands what he or she is expected to do and that no mistakes creep into the collection of saliva samples (e.g., using the tube for another time point for a given collection time; participants depleting their saliva glands because they continue to chew on a gum instead of taking it out after sample collection). Having participants collect saliva while they are completing a questionnaire or similar task helps keep them occupied during the collection phase, and also reduces the slight embarrassment associated with collecting saliva in front of an experimenter and other participants.

For the collection of saliva, we use skirted 50-ml tubes. The size of the tubes makes it easy for participants to drool directly into a tube; the skirt ensures that the tube can stand on the desk without the use of a holding rack; and the design of the tubes allows us to use them for storage, freezing, and centrifuging. To stimulate saliva flow, we use Trident Original Flavor chewing gum, which has been shown to yield the least bias in steroid levels as compared to unaided saliva flow (Dabbs, 1991; see also Shirtcliff, Granger, Schwartz, & Curran, 2001, for problems




PLEASE LOCATE THE PLASTIC COLLECTION TUBE NUMBERED "1".

While you are working on the first questionnaire, we would like to collect some saliva from you. This will be the first of six saliva samples.

To collect saliva, please chew on the **CHEWING GUM** that we put on your desk and spit the saliva that accumulates in your mouth into the PLASTIC COLLECTION **TUBE NUMBERED "1".**

Please continue to collect saliva in this way until the CLEAR SALIVA FLUID in the tube stands as high as the **BLACK MARKING** on the side of the tube. Then screw the lid back on tightly to prevent any saliva from evaporating or spilling.

AFTER YOU HAVE FILLED THE PLASTIC TUBE, PLEASE SPIT OUT THE CHEWING GUM.

If you have questions about this, please ask the experimenter at any time.

* Please press any key *

**FIGURE 3.2.** Sample computer instruction screen for the collection of saliva in a numbered 50-ml plastic tube.

associated with the use of cotton rolls for saliva collection). Before the start of data collection, all tubes had been properly labeled (using a waterproof permanent marker) with each participant's identification number; an additional number indicating the time point during the experiment at which the sample was taken (e.g., 101-1, 101-2, 101-3, 101-4, 101-5, 101-6); and a line at the 7.5-ml mark to indicate the desired fluid level. In our experience, a sample of this size can be collected quickly and provides enough material for the assessment of up to three different hormones.

After all samples had been collected in a session and participants had left the lab, the experimenter checked whether all tubes were properly sealed and put them into a freezer for storage. A regular chest freezer or upright freezer from a household appliance store is sufficient for this purpose. Frozen samples can last up to a year at –20°C and possibly longer without noticeable changes in the samples' steroid hormone concentration (Dabbs, 1991).

### *Sample Processing*

The goal of the second phase—saliva processing—is to make the saliva samples amenable to precise pipetting in the actual assay. To achieve this goal, all samples are first thawed and frozen three times after all data collection has been completed for the study. This procedure helps to break down the long molecule chains that make saliva sticky and viscous, and to turn it into a more watery (and thus precisely pipettable) fluid. The breakdown of molecular chains can be enhanced by speeding up freezing and thawing through the use of dry ice and a warm water bath; the stronger shearing forces associated with the fast temperature differential induced by the use of these aids facilitates the degradation of the molecule chains. After the third thaw, samples are spun for 10 min at 1000 g in a refrigerated centrifuge to push all coarse content to the bottom of the tube (this process is similar to the separation of serum and plasma in blood samples). After centrifugation, the "supernatant" (i.e., the watery part of the sample that stays on top after centrifugation) of each sample is transferred from the 50-ml tube to an identically labeled, or set of identically labeled, aliquot tube(s) (e.g., 5-ml, 2-ml, or 1.5-ml tubes). Care must be taken to avoid stirring up and transferring the coarse, sticky contents of saliva from the bottom of the tube during transfer. For this reason, we recommend centrifuging and aspirating only small batches of tubes (≤ 12) at a time, because coarse and watery components of saliva tend to mingle again after long waits between centrifugation and sample transfer to aliquots, particularly if samples are

not refrigerated during and after centrifugation. After aliquoting, samples can either be assayed right away or refrozen for later assaying.

### *Sample Assaying*

The goal of assaying the samples is to provide a specific, sensitive, accurate, and reliable measurement of their hormone content. Assay quality depends in part on the care taken during the previous steps. For instance, if participants have not been instructed to refrain from eating and brushing their teeth before the experiment, samples may be contaminated with traces of blood from gum lacerations and may therefore yield elevated steroid measurements, thus compromising accuracy (Granger et al., 2007). And if samples have not been processed thoroughly, saliva will be more likely to retain its viscous properties, which makes it difficult to pipette exact amounts of samples into test tubes and compromises measurement reliability.

Assay quality also depends on the properties of the assay itself. A key prerequisite for any assay is that it measures one hormone *specifically* (e.g., testosterone), but not other hormones (e.g., progesterone, estradiol, cortisol; see the first row of Table 3.1). The specificity requirement is sometimes difficult to satisfy, due to the close structural similarities of hormone molecules. An assay that claims to measure testosterone but in fact also responds to other androgens, such as androstenedione, may yield inflated estimates of the actual testosterone content of samples. It is often difficult to evaluate the specificity of a given assay, because running specificity tests in one's own lab requires work, time, and knowledge. But some telltale signs can be used to gauge whether a given assay indeed measures only what it claims to measure. If a sufficient number of subjects have been tested, one can compare the average hormone levels measured to those reported by others who have used high-pressure liquid chromatography (the gold standard in hormone assessment) and other assay methods to measure the same hormone in similar populations in saliva under similar circumstances (e.g., Dabbs et al., 1995). Thus, if one's measured hormone levels are in the same general range as hormone levels reported in the current literature (i.e., within 50–150% of the average levels measured by others), this can serve as an indicator of the assay's specificity. Another indicator is the detection of telltale circadian or group differences, particularly if their magnitude is similar to those observed by others. For instance, on average women typically have one-fourth to one-sixth of the testosterone levels measured in men at the same time of day. It should be a cause for concern if this difference is not observed in one's own samples with a given assay, provided

**TABLE 3.1. Overview of the Main Assay Quality Parameters**

| Assay quality parameter | Definition and estimation |
|---|---|
| Specificity | Defined as the ability of the assay to maximize measurement of the targeted analyte and minimize measurement of other analytes. Specificity is often established by measuring the degree to which an assay produces measurements different from 0 for nontargeted analytes (e.g., in the case of a cortisol assay, measurements > 0 for progesterone, aldosterone, pregnenolone, and other related steroid hormones). Cross-reactivity with such nontarget analytes is estimated by dividing the measured, apparent concentration of the analyte by the amount added. For example, if 2,000 µg/dL aldosterone is added, and 0.6 µg/dL is measured, (0.6/2000) • 100 = 0.03% cross-reactivity. Measures of specificity are not routinely included in hormone assays, but specificity should at least be carefully examined when a new assay is adopted. |
| Sensitivity | Defined as the lowest dose of an analyte that can be distinguished from a sample containing no analyte. It is often pragmatically derived by calculating the "lower limit of detection" (LLD), which is defined as signal obtained from a sample with zero analyte ($B_0$), minus three times the standard deviation of the signal at $B_0$. Values outside the $B_0 - (3 \times SD)$ range are considered valid nonzero measurements. |
| Accuracy | Defined as the ability of the assay to measure the true concentrations of the analyte in the samples being tested. Accuracy is measured by including control samples with known amounts of analyte in the assay and then comparing the amount of analyte estimated by the assay (e.g., 95 pg/ml) with the actual amount added (e.g., 100 pg/ml testosterone). The result is expressed as the percentage of the actual amount that is recovered by the assay; for example, accuracy = (95/100) · 100 = 95%. Recovery coefficients between 90% and 110% reflect good accuracy. |
| Precision | Defined as the closeness of agreement between test results repeatedly and independently obtained under stable conditions. Precision is typically estimated by the "coefficient of variation" (CV), which is calculated as the mean of replicate measurements of a given sample, divided by the standard deviation of the measurements, multiplied by 100. The intra-assay CV is calculated as the average of the CVs of all samples in a given assay or set of assays; the interassay CV is calculated from the between-assay mean and *SD* of a control sample (e.g., a saliva pool) included in all assays. Intra- and interassays CVs less than 10% are considered good. |

that enough subjects from each gender have been tested. Note that these indicators provide at best a rough, indirect test of an assay's ability to measure one specific hormone free of bias from other hormones, and do not constitute positive proof of an assay's specificity. Nonetheless, for a researcher whose interests and expertise lie not in biochemistry but in behavioral endocrinology, close examination of such indicators is imperative if the researcher wants to be able to trust in the validity of his or her hormone data.

Another important determinant of assay quality is the assay's *sensitivity* (see the second row of Table 3.1). Salivary assays for most major steroids are sufficiently sensitive to accurately discriminate even tiny differences—usually in the nanogram (i.e., 1/1,000,000,000 g) or even picogram (i.e., 1/1,000,000,000,000 g) range—in hormone concentrations across the entire range of concentrations usually observed in healthy adult populations. But some hormones and some populations push the limits of what many assays are able to detect and differentiate. For instance, estradiol is a powerful steroid hormone, subserving many different functions in the CNS and peripheral organs. Its enzymatic conversion from hormonal precursors (e.g., testosterone) is therefore tightly constrained, and overall levels in adults are low, with men and women in the follicular phase of the menstrual cycle exhibiting average salivary levels of 1–5 pg/ml. However, most estradiol assays reliably measure only levels of 2 pg/ml and up, and either cannot differentiate levels lower than 2 pg/ml from zero concentrations or yield unreliable or invalid measurements for concentrations below the 2-pg/ml threshold. Likewise, many steroid hormone levels are considerably lower in prepubertal children and in aging or menopausal populations. Assays that are sufficiently sensitive to cover the range of levels typically observed in healthy, fertile, adult populations may therefore not accurately measure the lower levels observed in other populations. Careful consideration of the typical hormone levels reported for such populations in the literature, and the subsequent selection of suitable assays that cover the lower range of hormone levels or the modification of various parameters (e.g., use of a preincubation phase, increase of sample added to assay) of previously used assays, are frequently necessary in such cases.

If an assay is sufficiently specific and sensitive, its *accuracy* over the entire range of measurements needs to be established (see the third row of Table 3.1). This essential validity check can be easily performed by adding control samples with known amounts of an analyte to the assay and monitoring the extent to which the measured concentration matches the expected concentration. Accuracy checks are usually done by including commercially available control samples (e.g., the Lyphochek samples

provided by BioRad, Hercules, CA). These frequently contain an array of different hormones and can thus be used for several different hormone assays. However, because they are calibrated to accommodate the hormone levels typically observed in blood, proper dilution is necessary for their use in salivary hormone assays. Accuracy checks should cover the low (25%), medium (50%), and high (75%) range of expected measurements for a given assay type—or, at a minimum, the lower and upper 33% of expected levels. Accuracy checks can and should also be included in batches of samples that are sent out to commercial assay services, and their inclusion should not be indicated to the assay company. They represent the only independent check of an assay's validity that a researcher can perform on the results returned by such companies.

Finally, measurement *reliability* (or *precision*) needs to be established and evaluated (see the fourth row of Table 3.1). This is usually done in two ways. First, all samples are assayed in duplicate, and the degree to which two measurements of the same sample differ, expressed as the "coefficient of variation" (CV), is averaged across all samples and reported as the intra-assay CV. Intra-assay CV is heavily influenced by the consistency of the assayer's pipetting technique; fluctuations in the way samples are pipetted into the assay tubes or plate wells, or improper handling of the pipette itself, increase the intra-assay CV. But other factors also play a role. In our lab, we have found preincubation of salivary testosterone and cortisol RIAs to yield lower CVs, presumably because bonds between salivary hormones and antibodies become more stable during preincubation. We have also observed that CVs for salivary steroid levels drop substantially if the source tube with the aliquoted sample is gently inverted two to three times before the sample is transferred to the assay. This observation suggests that steroid hormones can be unevenly distributed in the fluid column inside a sample tube after prolonged storage, perhaps due to the displacing force of other molecules that differ by size and density.

The intra-assay CV tends to exaggerate lack of reliability at the lower range of the assay, and thus for low-concentration samples, because the same absolute difference (e.g., 14 pg/ml vs. 18 pg/ml) between two duplicate measurements at lower concentrations yields a higher CV than at higher concentrations (e.g., 84 pg/ml vs. 88 pg/ml). It should be noted, too, that rank-order stability of sample concentrations, as estimated by Spearman's correlation coefficient or other measures of between-sample correlation, can be high despite an unsatisfactory intra-assay CV, because the variance between different individuals' hormone levels is frequently substantially higher than the measurement variance within samples (see Stanton & Schultheiss, 2007, for an example).

The second method of establishing measurement reliability is to determine interassay CV. With large batches of samples and many research participants, samples are usually split into several consecutive assays, each with its own set of calibrators for the generation of a standard curve. But if each assay gets its own measurement device (i.e., the calibrators), how can researchers ensure that different consecutive assays yield comparable readings? The solution to this problem is easy: including samples coming from the same source in all assays, and comparing readings for these samples across assays by determining the interassay CV. The sources should be sufficiently large to provide samples for all assays in a given series, and should ideally cover at least two substantially different analyte concentrations. In our own labs, we frequently either pool leftover samples or collect large quantities (> 10 ml) of saliva from lab members, and include these samples in all assays for a given study. To create pools with sufficiently different concentrations, we take known factors that affect hormone concentrations into consideration. For instance, because men and women differ greatly in their testosterone levels, we create separate male and female saliva pools. And to create pools with different cortisol concentrations, we separately pool samples collected early in the morning (when levels are high) and samples collected late at night (when cortisol is close to its circadian low point). When only a small number of samples have been collected that can be accommodated by one single assay for a given hormone, interassay CV cannot be determined. In this case, only the intra-assay CV is reported.

Researchers who would like to employ salivary hormone measures, but do not have the facilities to do the assays themselves, can "outsource" saliva sample analysis to commercial assay labs that specialize in salivary hormone measurement. However, in this case we strongly recommend that researchers not simply trust the claims these labs are making, but actually test their validity before and after sending off the samples. Of course, a thorough understanding of the quality parameters of good endocrine measurement as outlined above and in Table 3.1 is essential for this. One simple way to pick a good assay service is to compare the claims of the assay provider with the published literature. An assay service that uses an assay whose analytical range does not cover the expected hormone concentration range in the tested sample reasonably well (e.g., an estradiol range of 2–40 pg/ml, when average estradiol levels in men and women are typically about 1–4 pg/ml), or that reports excellent recovery coefficients for accuracy checks whose levels are far above the levels typically expected (for the estradiol example, at 10 and 25 pg/ml, which is substantially outside the range of values ordinarily observed in men or normally cycling women), should be viewed with some suspicion. As suggested previously, we also recommend includ-

ing one's own accuracy checks calibrated for the hormone concentrations expected in the study sample (e.g., cortisol accuracy checks at 1.5 and 3.5 ng/ml, corresponding to low and high salivary levels of this hormone). The investment in a set of commercially available calibrator samples (e.g., the Lyphochek samples mentioned above), a pipette, and a couple of tubes and pipette tips is comparatively minuscule (less than $700) and pays off in the form of an all-important independent verification of the quality of the outsourced assays. Finally, customers of commercial assay services should expect to receive a complete set of data that includes not only the mean hormone level and CV for each sample, but also the values for each individual measurement (for verification of the intra-assay CV); the values for standard pools used across assays (for verification of interassay CVs); and the complete data on the standard curve, including the zero-concentration calibrator, which can be used to verify the service's claims about the sensitivity of the assay. Schultheiss and colleagues (2005) checked assay specificity by comparing salivary testosterone concentrations to levels reported in the literature; determined and reported the "lower limit of detection" (LLD) as a measure of assay sensitivity (1 pg/ml); and also reported interassay CV (6.62%) and intra-assay CV (4.72%). The interassay CV estimate was based on an in-house saliva pool and three Lyphochek control samples, from which assay accuracy was also determined. Accuracy was excellent, with measured levels corresponding closely to expected levels for the low (59 pg/ml: 97%), medium (125 pg/ml: 101%), and high (250 pg/ml: 99%) range of male salivary testosterone levels. The analytical range (i.e., the range from the lowest to the highest nonzero standard) of the assay was 5–400 pg/ml and was sufficient to cover the observed levels of salivary testosterone, which ranged from 7 to 248 pg/ml.

### *Data Processing*

Regardless of whether one uses an RIA or EIA, the actual measurements returned by one's measurement device (be it a gamma counter or a plate reader) are not hormone concentrations but proxy measurements, such as counts per minute in the case of the gamma particles emitted by a decaying isotope. These need to be transformed and interpreted to make sense as hormone concentrations. Of course, many counters and readers come with built-in analysis software that can be programmed to automatically estimate analyte concentrations in samples. In our experience, however, this software is rarely useful for processing data resulting from salivary hormone analysis, because it is usually geared toward the determination of plasma hormone levels. And if it is employed without proper knowledge of the steps involved in assay data processing, and

without checking whether all prerequisites are met for the application of the software and the preset parameters, results can be severely biased or simply useless. We therefore advocate using one's preferred statistical software package for the mindful processing of hormone assay data. An excellent, hands-on guide to the steps involved in this process is provided by Nix and Wild (2000); here, we only give a brief overview of the main stepping-stones on the way from the raw measurements to the final concentration estimates.

The first step is usually to put the raw measurements in relation to the known concentrations of the calibrator samples with known concentrations, which constitute the standard curve (see Figure 3.3). This step frequently requires the computational transformation of the calibrator concentrations, the raw measurements, or both to bring calibrator concentrations and measurements into a linear relationship. It also requires the close examination of graphs depicting the relationship between known calibrator concentrations and observed measurements, to control for outliers that might bias the regression equation for the standard curve. In general, a regression of measured signal on calibrator concentrations is expected to exceed 97% of explained variance—a criterion that brings tears to the eyes of many a behavioral scientist, but represents an absolutely reasonable and defensible requirement for the deriva-

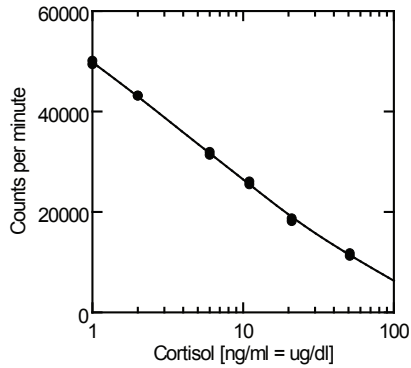



**FIGURE 3.3.** Standard curve from salivary cortisol assay. Standard concentrations cover the entire range of salivary cortisol levels typically observed in healthy populations (i.e., from 0.05 to 20 ng/ml). The standard with a concentration of 0 ng/ml is not depicted here, but is included in the assay. A distance-weighted least-squares regression line is fitted through the data points. Note that log-transformation of the *x* axis makes the relationship between hormone concentration (*x* axis) and gamma emissions (in counts per minute; *y* axis) approximately linear.

tion of the regression equation that will be used to accurately estimate concentrations in unknown samples. In our experience, including three or more samples for each calibrator concentration greatly facilitates the determination of a robust standard curve, because average concentrations at each level are less likely to be influenced by outliers.

Once a good, linear fit between hormone concentration in the calibrators and measured signal (e.g., counts per minute) has been determined, the relationship between predictor and dependent variable is turned upside down to move from the question "What signal level does a given concentration predict?" to "What concentration can I infer, given a certain signal level?" The flipping of the relationship between predictor and dependent variable entailed in this second step is simplified if the relationship between both variables has been linearized previously, and the regression equation thus becomes symmetrical with regard to the predictor and the criterion. The prediction of calibrator concentrations from the measured signal yields a regression equation that can also be applied to all other, noncalibrator samples—that is, to control samples and samples collected from the research participants.

The application of this formula to the rest of the samples in the assay constitutes the third step. Because the variance around the zero-concentration calibrator and the mean signal levels of control samples can be interpreted in terms of actual analyte concentrations, it is also possible to determine the sensitivity and the accuracy, respectively, of the assay at this point.

In a final step, after data from the calibrators and the control samples have been discarded, the CV of the duplicate measurements for the actual samples collected from research participants is determined, and mean concentration levels for each sample are calculated. These mean estimated concentrations are then used as interval-scale variables in subsequent analyses (regression and correlation analyses, analyses of variance [ANOVAs], analyses of covariance [ANCOVAs], etc.) probing the validity of the research hypothesis.

In the case of the analyses for Study 1 in Schultheiss and colleagues (2005), a repeated-measures ANCOVA design was used. The researchers entered testosterone levels measured immediately before the contest start (T3) as a covariate, postcontest testosterone (T4, T5, T6) as a within-subjects variable, and contest outcome (win vs. lose) and implicit power motivation (*z* scores) as between-subjects predictors. Schultheiss and colleagues found a power motive × contest outcome × time effect that was consistent with hypotheses: 15 min after the contest (T5) but not earlier (T3) or later (T6), power motivation significantly predicted testosterone increases in winners and decreases in losers (see Figure 3.4).

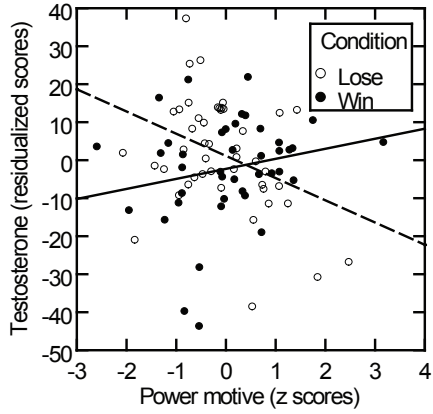



**FIGURE 3.4.** Effect of dispositional power motivation (*z* scores, *x* axis) and experimentally varied contest outcome (win, lose) on male participants' testosterone levels 15 min after the contest (T5; *y* axis). Postcontest testosterone was residualized for testosterone levels immediately before the contest (T3), which served as baseline. Partial correlations for the effect of power motivation on testosterone changes are .21 for winners and –.38 for losers. From Schultheiss et al. (2005). Copyright 2005 by the American Psychological Association. Reprinted by permission.

### *Reporting the Results of a Hormone Assay*

As the reporting of other findings does, reporting the results of hormone assays involves two steps. First, the method of assessment and its quality should be reported in the "Methods" section. Second, the actual findings are reported in the "Results" section.

Description of the method should include the exact type and make of the assay; a short summary of the sample processing and sample assay protocol, particularly of points where they diverged from routine protocols (e.g., if samples were pretreated in some way or a preincubation period was used); and also the main quality control parameters of the assay—that is, measures of validity (specificity, accuracy, LLD, analytical range) and reliability (intra- and interassay CV). For well-established assay procedures, it is usually sufficient to omit estimates of specificity and accuracy, and to report only analytical range, LLD, and CV. Assay quality parameters provided by the manufacturers of commercially available assays should *not* be reported, as these typically represent best-case scenarios that are included with the assay to promote sales and that may have little to do with the quality of an assay actually conducted in one's own lab.

Reporting of findings should include descriptive data on the hormone levels observed in the sample and their relationship to major influ-

ences on endocrine function, such as gender, menstrual cycle stage, use of oral contraceptives, and time of day when samples were collected. In general, the same rules and best practices for analyzing and reporting other kinds of data also apply to hormone measures. Thus hormone data distributions should be examined for skew and, if necessary, transformed to bring them closer to a normal distribution (this is frequently necessary for salivary cortisol data and may be required for other hormones in some cases); if this is done, it should be reported. If outliers are present in the hormone data (e.g., elevated estradiol due to ovulation, high progesterone levels sometimes observed in women in the luteal phase or in the early stages of pregnancy, or extreme levels of cortisol sometimes observed in individuals with undiagnosed endocrine disorders), and they cannot be accommodated through standard data transformations, analyses should be run and reported both with and without the outliers. If the findings hold up to scrutiny either way, nothing is lost by pointing this out; if they emerge only in one or the other case, this needs to be considered in the "Discussion" section and perhaps even before the paper is written.

## Advantages and Disadvantages of Salivary Hormone Measures

Toward the end of this chapter's introduction, we have already pointed out that a main advantage of salivary hormone measures is that they are pain-free and thus easy to use in behavioral studies, and that their main disadvantage is the limitation of their use to those hormones that make it into saliva. With regard to the latter issue, it is important to keep in mind that the method should never dictate the research question, and that if it is conceptually reasonable to assess a hormone that is not present in saliva but in other body compartments, a researcher should consider alternative methods. For instance, metabolites of some peptide hormones can be assessed in urine; peptide hormone levels can be varied experimentally through nasal administration of sprays (e.g., Born et al., 2002); and if a hormone of interest can only be assessed in blood, then the researcher could team up with a physician, nurse, or phlebotomist to get the necessary blood samples from participants and take proper care to minimize the effects of venipuncture-induced stress on the measurement of the targeted hormone (e.g., by allowing a sufficient amount of time after venipuncture for the participant to relax and get used to the measurement situation). In this section, we briefly touch on some conceptual advantages and disadvantages of hormone assessment that we think are important for researchers interested in adding endocrine measures to their armamentarium.

A major advantage of salivary hormone measures is that they simultaneously meet personality psychologists' need for rank-order stability and social psychologists' need for measures that are sensitive to the social stimuli impinging on the person. In Study 1 of the data reported by Schultheiss and colleagues (2005), correlations of testosterone concentrations in consecutive saliva samples ranged from .82 to .96, indicating high within-session stability of male testosterone levels. For a testing interval of 48 hours, Sellers, Mehl, and Josephs (2007) recently reported salivary testosterone test–retest correlations of .69 for men and .72 for women. Even for saliva samples taken 2 months apart, Dabbs (1990) reported test–retest correlations between .43 and .59. Comparable findings have been reported for salivary estradiol (Stanton & Schultheiss, 2007) and cortisol (Cieslak, Frost, & Klentrou, 2003). Within-session stability of salivary progesterone levels was high (.80 to .87) in the Schultheiss and colleagues (2004) study, and current investigations in our laboratories suggest that salivary progesterone also shows substantial retest stability over the course of 2 weeks.

Stable differences in salivary hormones in turn are associated with important behavioral differences. Individual differences in salivary testosterone correlate positively with measures of dominant and aggressive behavior (e.g., Dabbs & Hargrove, 1997; Dabbs, Jurkovic, & Frady, 1991; Schaal, Tremblay, Soussigan, & Susman, 1996) and predict attention and learning in response to facial expressions (van Honk et al., 2000; Wirth & Schultheiss, 2007). Likewise, individual differences in salivary cortisol levels are predictive of attentional responding to threat stimuli (van Honk et al., 1998) and can interact with testosterone in shaping behavior (e.g., Dabbs et al., 1991).

In this context, it is remarkable that self-report measures of personality and emotionality notoriously fail to correlate substantially with basal measures of steroid hormones. For instance, salivary and other measures of testosterone have no consistent variance overlap with questionnaire measures of dominance and aggression (Mazur & Booth, 1998). Moreover, a meta-analysis revealed that cortisol, despite its reputation as a "stress hormone," does not correlate with questionnaire measures of negative emotionality (Dickerson & Kemeny, 2004). At the same time, the roles of these hormones in dominance and stress responses are well documented in humans and in a wide variety of nonhuman and even nonmammalian species (see Nelson, 2005), suggesting considerable phylogenetic continuity of the functions these hormones fulfill. Also, the mechanisms through which hormones affect behavior and are affected by situations have been worked out in great detail in many cases (e.g., Albert, Jonik, & Walsh, 1992; Sapolsky, 1987; Schultheiss, 2007a). In contrast, we humans are the only species capable of filling out ques-

tionnaires inquiring about our dispositions to dominate, affiliate, experience negative affect, and so on, and the extent to which the beliefs measured through such instruments can be mapped onto the embrained and embodied systems that guide actual dominance behavior, affiliation, stress responses, and so on is often unclear (cf. Gazzaniga, 1985; Kagan, 2002; McClelland, Koestner, & Weinberger, 1989; Schultheiss, 2007b). Therefore, hormonal measures of dominance and other behavioral dispositions have the clear advantage of being more likely to carve nature at its joints than questionnaire measures of the same constructs.

Salivary hormone measures are also sufficiently sensitive to situational stimuli and events to be of use to social psychologists. Examples of the effects of the situation on people's salivary hormone levels include research on the impact of affiliation-arousing movies on progesterone and cortisol (Schultheiss et al., 2004; Wirth & Schultheiss, 2006); competition effects on testosterone (e.g., Gladue et al., 1989; Mazur, Booth, & Dabbs, 1992); and, most thoroughly documented, effects of social-evaluative threat on cortisol changes (Kirschbaum, Pirke, & Hellhammer, 1993; see also the meta-analysis by Dickerson & Kemeny, 2004).

In our view, however, salivary hormone measures are most usefully employed in research that combines dispositional with situational factors, because individuals bring their learning history, genetic makeup, and so forth to a given situation and respond to the situation on the basis of these dispositional factors (cf. also Sapolsky, 1999). Our own worked-out example above has already illustrated the interactive effect of implicit power motivation (disposition) and contest outcome (situation) on testosterone changes. Notably, in this study as in many others, contest outcome per se did not have a main effect on hormonal changes, because the hormonal responses to competitions in individuals who do not seek power are virtually the opposite of the responses in those who do (see also Josephs, Sellers, Newman, & Mehta, 2006; Wirth, Welsh, & Schultheiss, 2006). Work by Josephs and colleagues (2006) illustrates that salivary hormone measures can also be used as predictors, rather than as dependent variables, in person × situation designs. Based on the hypothesis that high levels of salivary testosterone should make people feel comfortable when in a high-power position but uncomfortable in a low-power position, and that low levels of testosterone should be associated with the reverse, Josephs and colleagues showed that low-testosterone participants reported greater emotional arousal and showed worse cognitive functioning in a high-power position, whereas high-power individuals exhibited this pattern in a low-power position. Giving the person × situation design a further twist, Mehta and Josephs (2006) recently demonstrated that men who responded with a salivary testosterone *decrease* to a defeat in a competition were less likely to enter another

dominance contest than were men who responded with a testosterone *increase*.

One potential drawback of the use of endocrine measures in social neuroscience is that hormones do not map onto established psychological constructs in a one-to-one fashion, or do so only within certain boundaries. For instance, cortisol increases in response to a public speaking task are readily interpreted as a physiological indicator of stress. But what then about the fact that cortisol also surges after a meal? Or that cortisol reaches its highest peak in the morning, just before people get up? Clearly, cortisol overlaps only partially with the concept of stress, and the energy-regulating functions of cortisol encompass more than just dealing with a stressful situation. As another case in point, high or increasing levels of testosterone are associated with aggression and dominance. But testosterone also regulates libido, energy metabolism, and tissue buildup in the muscles, and many of the central (i.e., brain) effects of testosterone are mediated fully or in part by its powerful metabolite estradiol. Furthermore, the numerous interactions of hormonal systems with each other, as well as with the immune system, brain systems, and peripheral organs, are often complex and intimidating at first glance. The only way to master this complexity and to employ hormone measures successfully in one's research requires a thorough understanding of the endocrine system.

We believe, however, that these apparent disadvantages can be turned into advantages if one is willing (1) to become acquainted with the basic literature on endocrine systems (e.g., Griffin & Ojeda, 2000), their relationships with brain and behavior (e.g., Nelson, 2005), and their assessment (e.g., Gosling, 2000; Riad-Fahmy, Read, Walker, & Griffiths, 1982; Riad-Fahmy et al., 1987); and (2) to be curious about the ramifications of casting one's research hypothesis in endocrinological terms, and to keep an open mind about the findings one obtains from research employing hormonal measures. The apparent disadvantage of complex interactions between hormonal and other systems can quickly turn into an advantage if one realizes that each hormone comes with a rich, multidisciplinary research literature, and that even after relatively coarse perusal this literature can suggest exciting new hypotheses and research directions that would not have been apparent otherwise. For instance, recent research that builds on a knowledge of the organizational effects of testosterone on male facial morphology, and the activational effects of estradiol on female mate choice, demonstrates that women with higher estradiol concentrations exhibit stronger preferences for the faces of men with higher testosterone concentrations—an effect that changes with menstrual cycle stage and thus with circulating estradiol levels (Roney & Simmons, 2008). Because some of the effects of hor-

mones on brain systems mediating specific cognitive functions have been worked out in great detail in animal models, it is no longer far-fetched to postulate and study hormone-mediated effects of stress and social factors on memory and other cognitive processes in humans (e.g., Cahill, 2000; Putman, van Honk, Kessels, Mulder, & Koppeschaar, 2004). And the interface among social psychology, endocrinology, and immunology holds particular promise for a better understanding of how experiential factors affect physical health and illness (e.g., Munck, Guyre, & Holbrook, 1984; Sapolsky, 2004).

In summary, incorporating endocrine measures and concepts in one's research requires work, as all efforts at broadening one's horizon do. But speaking from experience, we believe that it is work well invested, because interfacing psychology with endocrinology makes the study of human behavior more rigorous, intellectually stimulating, and (perhaps most importantly) likely to yield exciting discoveries. And *that* gets our hormones going!

## APPENDIX 3.1. PC-Administered Screening Questionnaire Used in Studies with Hormone Assessments

(Time of day is recorded by the PC.)
Please enter your age.
Please enter your gender.
Please enter your weight (in pounds).
Please enter your height.
Have you experienced any gum bleeding over the past day?
Have you experienced any other oral infections and/or oral lacerations over the past day?
How long ago, in hours, has it been since you brushed your teeth?
How many hours ago has it been since you consumed caffeine (coffee, tea, soda, chocolate)?
How many hours ago has it been since you consumed an alcoholic beverage?
Are you currently on any kind of medication? If yes, please provide the name of the prescription.
Do you have a diagnosed endocrine disorder? If yes, please name the disorder.
Do you use any recreational drugs (e.g., marijuana, Ecstasy, speed, cocaine, heroin)?
Do you smoke?
Do you take anabolic steroids?
Are your currently involved in a steady relationship?
Have you had sexual intercourse in the last 24 hours?
Please indicate the hand (left or right) you typically use in activities such as writing, brushing your teeth, holding a glass, etc.

Additional questions for women only:
What was the date on which your last menstrual period started?
What is the average duration of your menstrual cycle (in days)? (By "menstrual cycle," we mean the time from the start of one menstrual period to the start of the next.)
Do you currently take oral contraceptives (i.e., the "pill")?

## Acknowledgment

We thank Michelle M. Wirth for her constructive comments and valuable suggestions on an earlier version of this chapter.

## References

Recommended introductory readings are marked with an asterisk (*).

Albert, D. J., Jonik, R. H., & Walsh, M. L. (1992). Hormone-dependent aggression in male and female rats: Experiential, hormonal, and neural foundations. *Neuroscience and Biobehavioral Reviews, 16*(2), 177–192.

Baron-Cohen, S., Lutchmaya, S., & Knickmeyer, R. (2006). *Prenatal testosterone in mind: Amniotic fluid studies*. Cambridge, MA: MIT Press.

Beckwith, B. E., Petros, T. V., Bergloff, P. J., & Staebler, R. J. (1987). Vasopressin analogue (DDAVP) facilitates recall of narrative prose. *Behavioral Neuroscience, 101*(3), 429–432.

Born, J., Lange, T., Kern, W., McGregor, G. P., Bickel, U., & Fehm, H. L. (2002). Sniffing neuropeptides: A transnasal approach to the human brain. *Nature Neuroscience, 5*(6), 514–516.

Cahill, L. (2000). Modulation of long-term memory in humans by emotional arousal: Adrenergic activation and the amygdala. In J. P. Aggleton (Ed.), *The amygdala: A functional analysis* (pp. 425–446). New York: Oxford University Press.

Carter, C. S., Pournajafi-Nazarloo, H., Kramer, K. M., Ziegler, T. E., White-Traut, R., Bello, D., et al. (2007). Oxytocin: Behavioral associations and potential as a salivary biomarker. *Annals of the New York Academy of Sciences, 1098*, 312–322.

Cieslak, T. J., Frost, G., & Klentrou, P. (2003). Effects of physical activity, body fat, and salivary cortisol on mucosal immunity in children. *Journal of Applied Physiology, 95*(6), 2315–2320.

Dabbs, J. M. (1990). Salivary testosterone measurements: Reliability across hours, days, and weeks. *Physiology and Behavior, 48*, 83–86.

Dabbs, J. M. (1991). Salivary testosterone measurements: Collecting, storing, and mailing saliva samples. *Physiology and Behavior, 49*, 815–817.(*)

Dabbs, J. M. (1992). Testosterone measurements in social and clinical psychology. *Journal of Social and Clinical Psychology, 11*, 302–321.(*)

Dabbs, J. M., Jr., Campbell, B. C., Gladue, B. A., Midgley, A. R., Navarro, M.

A., Read, G. F., et al. (1995). Reliability of salivary testosterone measurements: A multicenter evaluation. *Clinical Chemistry, 41*(11), 1581–1584.
Dabbs, J. M., & Hargrove, M. F. (1997). Age, testosterone, and behavior among female prison inmates. *Psychosomatic Medicine, 59*, 477–480.
Dabbs, J. M., Jurkovic, G. J., & Frady, R. L. (1991). Salivary testosterone and cortisol among late adolescent male offenders. *Journal of Abnormal Child Psychology, 19*, 469–478.
Daly, R. C., Su, T. P., Schmidt, P. J., Pagliaro, M., Pickar, D., & Rubinow, D. R. (2003). Neuroendocrine and behavioral effects of high-dose anabolic steroid administration in male normal volunteers. *Psychoneuroendocrinology, 28*(3), 317–331.
Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*(3), 355–391.
Gazzaniga, M. S. (1985). *The social brain: Discovering the networks of the mind*. New York: Basic Books.
Gladue, B. A., Boechler, M., & McCaul, K. D. (1989). Hormonal response to competition in human males. *Aggressive Behavior, 15*, 409–422.
Gosling, J. P. (Ed.). (2000). *Immunoassays: A practical approach*. Oxford, UK: Oxford University Press.(*)
Graham, J. M., & Desjardins, C. (1980). Classical conditioning: Induction of luteinizing hormone and testosterone secretion in anticipation of sexual activity. *Science, 210*, 1039–1041.
Granger, D. A., Cicchetti, D., Rogosch, F. A., Hibel, L. C., Teisl, M., & Flores, E. (2007). Blood contamination in children's saliva: Prevalence, stability, and impact on the measurement of salivary cortisol, testosterone, and dehydroepiandrosterone. *Psychoneuroendocrinology, 32*(6), 724–733.
Griffin, J. E., & Ojeda, S. R. (Eds.). (2000). *Textbook of endocrine physiology* (4th ed.). Oxford, UK: Oxford University Press.(*)
Hibel, L. C., Granger, D. A., Kivlighan, K. T., & Blair, C. (2006). Individual differences in salivary cortisol: Associations with common over-the-counter and prescription medication status in infants and their mothers. *Hormones and Behavior, 50*(2), 293–300.
Hofman, L. F. (2001). Human saliva as a diagnostic specimen. *Journal of Nutrition, 131*(5), 1621S–1625S.
Horvat-Gordon, M., Granger, D. A., Schwartz, E. B., Nelson, V. J., & Kivlighan, K. T. (2005). Oxytocin is not a valid biomarker when measured in saliva by immunoassay. *Physiology and Behavior, 84*(3), 445–448.
Josephs, R. A., Sellers, J. G., Newman, M. L., & Mehta, P. H. (2006). The mismatch effect: When testosterone and status are at odds. *Journal of Personality and Social Psychology, 90*(6), 999–1013.
Kagan, J. (2002). *Surprise, uncertainty, and mental structures*. Cambridge, MA: Harvard University Press.
Kirschbaum, C., & Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. *Psychoneuroendocrinology, 19*(4), 313–333.(*)
Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The "Trier Social

Stress Test": A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*(1–2), 76–81.

Laeng, B., & Falkenberg, L. (2007). Women's pupillary responses to sexually significant others during the hormonal cycle. *Hormones and Behavior, 52*(4), 520–530.

Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Social Forces, 64*, 377–402.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences, 21*, 353–397.

Mazur, A., Booth, A., & Dabbs, J. M. (1992). Testosterone and chess competition. *Social Psychology Quarterly, 55*, 70–77.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96*, 690–702.

Mehta, P. H., & Josephs, R. A. (2006). Testosterone change after losing predicts the decision to compete again. *Hormones and Behavior, 50*(5), 684–692.

Munck, A., Guyre, P. M., & Holbrook, N. J. (1984). Physiological functions of glucocorticoids in stress and their relation to pharmacological actions. *Endocrine Reviews, 5*, 25–44.

Nadeem, H. S., Attenburrow, M. J., & Cowen, P. J. (2004). Comparison of the effects of citalopram and escitalopram on 5-HT-mediated neuroendocrine responses. *Neuropsychopharmacology, 29*(9), 1699–1703.

Nelson, R. J. (2005). *An introduction to behavioral endocrinology* (3rd ed.). Sunderland, MA: Sinauer Associates.(*)

Nix, B., & Wild, D. (2000). Data processing. In J. P. Gosling (Ed.), *Immunoassays: A practical approach* (pp. 239–261). Oxford, UK: Oxford University Press.

Oyegbile, T. O., & Marler, C. A. (2005). Winning fights elevates testosterone levels in California mice and enhances future ability to win fights. *Hormones and Behavior, 48*(3), 259–267.

Pope, H. G., Jr., Kouri, E. M., & Hudson, J. I. (2000). Effects of supraphysiologic doses of testosterone on mood and aggression in normal men: A randomized controlled trial. *Archives of General Psychiatry, 57*(2), 133–140 (discussion, 155–156).

Putman, P., van Honk, J., Kessels, R. P., Mulder, M., & Koppeschaar, H. P. (2004). Salivary cortisol and short- and long-term memory for emotional faces in healthy young women. *Psychoneuroendocrinology, 29*(7), 953–960.

Raff, H., Homar, P. J., & Burns, E. A. (2002). Comparison of two methods for measuring salivary cortisol. *Clinical Chemistry, 48*(1), 207–208.

Riad-Fahmy, D., Read, G. F., Walker, R. F., & Griffiths, K. (1982). Steroids in saliva for assessing endocrine function. *Endocrine Reviews, 3*(4), 367–395.(*)

Riad-Fahmy, D., Read, G. F., Walker, R. F., Walker, S. M., & Griffiths, K. (1987). Determination of ovarian steroid hormone levels in saliva: An overview. *Journal of Reproductive Medicine, 32*(4), 254–272.(*)

Roney, J. R., Lukaszewski, A. W., & Simmons, Z. L. (2007). Rapid endocrine

responses of young men to social interactions with young women. *Hormones and Behavior, 52*(3), 326–333.

Roney, J. R., & Simmons, Z. L. (2008). Women's estradiol predicts preference for facial cues of men's testosterone. *Hormones and Behavior*, *53*(1), 14–19.

Sapolsky, R. M. (1987). Stress, social status, and reproductive physiology in free-living baboons. In D. Crews (Ed.), *Psychobiology and reproductive behavior: An evolutionary perspective* (pp. 291–322). Englewood Cliffs, NJ: Prentice-Hall.

Sapolsky, R. M. (1999). Hormonal correlates of personality and social contexts: From nonhuman to human primates. In C. Panter-Brick & C. Worthman (Eds.), *Hormones, health, and behaviour: A socio-ecological and lifespan perspective* (pp. 18–46). Cambridge, UK: Cambridge University Press.

Sapolsky, R. M. (2004). *Why zebras don't get ulcers* (3rd ed.). New York: Holt.

Schaal, B., Tremblay, R. E., Soussigan, R., & Susman, E. J. (1996). Male testosterone linked to high social dominance but low physical aggression in early adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry, 34*, 1322–1330.

Schultheiss, O. C. (2007a). A biobehavioral model of implicit power motivation arousal, reward and frustration. In E. Harmon-Jones & P. Winkielman (Eds.), *Social neuroscience: Integrating biological and psychological explanations of social behavior* (pp. 176–196). New York: Guilford Press.

Schultheiss, O. C. (2007b). A memory-systems approach to the classification of personality tests: Comment on Meyer and Kurtz (2006). *Journal of Personality Assessment, 89*(2), 197–201.

Schultheiss, O. C., & Brunstein, J. C. (1999). Goal imagery: Bridging the gap between implicit motives and explicit goals. *Journal of Personality, 67*, 1–38.

Schultheiss, O. C., Campbell, K. L., & McClelland, D. C. (1999). Implicit power motivation moderates men's testosterone responses to imagined and real dominance success. *Hormones and Behavior, 36*(3), 234–241.

Schultheiss, O. C., Wirth, M. M., & Stanton, S. J. (2004). Effects of affiliation and power motivation arousal on salivary progesterone and testosterone. *Hormones and Behavior, 46*(5), 592–599.

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*(1), 174–188.

Sellers, J. G., Mehl, M. R., & Josephs, R. A. (2007). Hormones and personality: Testosterone as a marker of individual differences. *Journal of Research in Personality, 41*, 126–138.

Shirtcliff, E. A., Granger, D. A., Schwartz, E., & Curran, M. J. (2001). Use of salivary biomarkers in biobehavioral research: Cotton-based sample collection methods can interfere with salivary immunoassay results. *Psychoneuroendocrinology, 26*(2), 165–173.

Stanton, S. J., & Schultheiss, O. C. (2007). Basal and dynamic relationships between implicit power motivation and estradiol in women. *Hormones and Behavior, 52*(5), 571–580.

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*(1), 238–242.

van Honk, J., Tuiten, A., van den Hout, M., Koppeschaar, H., Thijssen, J., de Haan, E., et al. (1998). Baseline salivary cortisol levels and preconscious selective attention for threat: A pilot study. *Psychoneuroendocrinology, 23*(7), 741–747.

van Honk, J., Tuiten, A., van den Hout, M., Koppeschaar, H., Thijssen, J., de Haan, E., et al. (2000). Conscious and preconscious selective attention to social threat: Different neuroendocrine response patterns. *Psychoneuroendocrinology, 25*(6), 577–591.

Wirth, M. M., & Schultheiss, O. C. (2006). Effects of affiliation arousal (hope of closeness) and affiliation stress (fear of rejection) on progesterone and cortisol. *Hormones and Behavior, 50*, 786–795.

Wirth, M. M., & Schultheiss, O. C. (2007). Basal testosterone moderates responses to anger faces in humans. *Physiology and Behavior, 90*(2–3), 496–505.

Wirth, M. M., Welsh, K. M., & Schultheiss, O. C. (2006). Salivary cortisol changes in humans after winning or losing a dominance contest depend on implicit power motivation. *Hormones and Behavior, 49*(3), 346–352.