Measuring implicit motives

Oliver C. Schultheiss and Joyce S. Pang

University of Michigan, Ann Arbor

Appeared as: Schultheiss, O. C., & Pang, J. S. (2007). Measuring implicit motives. In R. W. Robins, R. C. Fraley & R. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 322-344). New York: Guilford.

Author note:

All picture stimuli described in this chapter are available upon request from the authors.

Correspondence concerning this chapter should be addressed to Oliver C. Schultheiss or Joyce S. Pang (oliver.schultheiss@psy.phil.uni-erlangen.de or joycepang@ntu.edu.sg).

Measuring implicit motives

Over the past ten years, the field of personality psychology has seen a resurgence of interest in implicit motives, nonconscious motivational needs that orient, select, and energize behavior (McClelland, 1987). This renewed interest has led to significant advances in our understanding of the biological, cognitive, emotional, social, and cultural manifestations of implicit motivational needs (e.g., Brunstein, Schultheiss, & Grässmann, 1998; Hofer & Chasiotis, 2003; Langner & Winter, 2001; Schultheiss, Pang, Torges, Wirth, & Treynor, 2005; Schultheiss, Wirth & Stanton, 2004; Schultheiss, Wirth, Torges, Pang, Villacorta & Welsh, 2005; Woike, 1995; Zurbriggen, 2000). Some strides have also been made towards conceptual refinements of the implicit motive construct (e.g., Brunstein & Maier, in press; Schultheiss, 2001a; Schultheiss & Brunstein, 2005).

However, as Schultheiss and Brunstein (2001) and Pang and Schultheiss (in press) recently pointed out, considerably less effort has gone into the description, evaluation, and refinement of the implicit motive measures commonly used in research. As a result, the thematic content coding method of motive assessment has remained an enigma to researchers not familiar with the field, and the threshold for using these methods may therefore be higher than for other, more frequently used assessment tools in personality psychology. The present chapter is intended to lower this threshold by (I) discussing the differential validity of implicit motive measures, as contrasted with measures of individuals' explicit, self-attributed motivational needs and goals, (II) reviewing the reliability of implicit motive measures, (III) providing a step-by-step guide for the assessment of implicit motives and the use of motive scores in analytic designs, (IV) illustrating the use of implicit motive measures in research with case examples, and (V) pointing out some future directions for the development of better motive measures.

I. Differential validity of implicit motive measures

Research on implicit motive constructs originated in the assumption that people do not have introspective access to many of the wellsprings of their behavior (McClelland, Atkinson, Clark & Lowell, 1953; see also Gazzaniga, 1985; Nisbett & Wilson, 1977; Kagan, 2002). McClelland, Atkinson and their colleagues therefore used motivation arousal studies to develop content coding measures of imaginative stories that research participants write in response to picture cues, a technique called the Picture Story Exercise (or PSE; cf. McClelland, Koestner, & Weinberger, 1989; see Winter, 1999, for key differences between the PSE and Murray's, 1943, Thematic Apperception Test). These content coding measures were designed to capture individuals' implicit needs for power (or n Power; a concern for having impact on others), achievement (or n Achievement; a concern for doing well according to a standard of excellence), affiliation (or n Affiliation; a concern for establishing, maintaining, or restoring close, friendly relationships), and intimacy (or n Intimacy; a preference for experiences of warm, close, and communicative interactions with others) (for a compilation of these and other scoring systems, see Smith, 1992a). Of course, it is also possible to ask people directly to what extent they view themselves as being motivated to have power, achieve, or affiliate with and be close to others, and over time many carefully constructed instruments were developed for this purpose, such as Jackson's (1984) Personality Research Form (PRF) or, more recently, goal inventories that assess people's commitments to specific power, achievement, affiliation, and intimacy goals in their daily lives (e.g., Brunstein et al., 1998).

The parallel development of thematic-content and self-report measures of motivation revealed a pervasive lack of variance overlap between the two types of assessment, which soon instigated a fight over which type of measure is the more valid one (e.g., Atkinson, 1981;

Entwisle, 1972; Fineman, 1977; McClelland, 1987). Eventually, this quarrel prompted researchers to delineate more carefully the differential validity of thematic-content measures of motivation (cf. McClelland et al., 1989; Schultheiss, 2001a; Spangler, 1992). Summarizing the main conceptual and empirical outcomes of this endeavor, Schultheiss (2001a, in press; Schultheiss & Brunstein, 2005) characterized the key characteristics of implicit motives as follows:

Implicit motives are more likely to become aroused by and respond to <u>nonverbal cues</u> than verbal stimuli (Schultheiss, 2001). Klinger (1967) observed that individuals responded with increases in affiliation or achievement motivation expressed in imaginative stories to watching an affiliation-oriented or achievement-oriented experimenter, even if they could not hear his verbal instructions. In a similar vein, Schultheiss and Brunstein (1999; 2002) demonstrated that experimenters who assigned and described a power-related goal verbally to their participants failed to arouse participants' power motive. Only after participants had had an opportunity to translate the verbally assigned goal into an experiential format through a goal imagery exercise did their power motive predict goal commitment and task performance. Finally, recent research indicates that facial expressions of emotion are particularly salient nonverbal cues for implicit motives. Facial signals of friendliness and hostility interact with individuals' implicit affiliation motive and facial signals of dominance and submission interact with individuals' implicit power motive to shape attentional orienting and instrumental learning (Schultheiss & Hale, submitted; Schultheiss, Pang, et al, 2005).

Consistent with their bias towards responding to stimuli that are not verbally encoded, implicit motives are particularly likely to show an effect on behavior if <u>non-declarative measures</u> (e.g., measures of behaviors and processes that are neither controlled by a person's conscious intentions nor the self-concept) are employed, but have very limited or no effects on declarative measures of motivation (i.e., measures that tap into a person's conscious sense of self and the beliefs, judgments, decisions, and attitudes associated with it). The differential effect of implicit motives on declarative and non-declarative measures was first documented by deCharms, Morrison, Reitman and McClelland (1955), who found that n Achievement predicted performance on a scrambled-word test (a non-declarative measure), but not participants' attribution of achievement-related traits to themselves or others (declarative measures of motivation). Later, Biernat (1989) showed that the implicit achievement motive predicted performance on a math task (non-declarative measure), but not subjects' decision to serve as a group leader on another task (declarative measure). In a similar vein, Brunstein and Hoyer (2002) found that high-achievement individuals showed superior performance on a vigilance task (non-declarative measure), but were not more likely than low-achievement individuals to continue on the task if given the choice (declarative measure). Lastly, Schultheiss and Brunstein (2002) found that the implicit power motive predicts nonverbal (e.g., gesturing, facial expressions) and paraverbal (e.g., speech fluency) behaviors on a persuasion task, but not the actual content of the arguments presented, which can be conceived of as a declarative measure.

To summarize, implicit motives are not accessible to introspection and self-report, respond preferentially to nonverbal stimuli, and predict non-declarative measures of motivation and behavior better than declarative measures. In addition, it is important to keep in mind that individual motives respond specifically to certain types of incentives, but not others. For instance, power-motivated individuals respond to opportunities to have impact on others, but not to situations that allow relaxed, friendly contact with others, whereas the reverse is true of affiliation-motivated individuals.

II. Reliability of implicit motive measures

The reliability of implicit motive measures has most frequently been estimated through (a) their internal consistency, (b), interscorer agreement, and (c) their test-retest reliability. PSE motive measures appear to have low reliability when their internal consistency is assessed, that is, when motive scores derived from stories written in response to different picture cues are correlated with each other. Typically, these correlations tend to be low, resulting in internal consistency coefficients (e.g., Cronbach's alpha) in the .20 to .50 range.

However, as Atkinson (1981; Atkinson & Birch, 1970) has pointed out in the context of his dynamics of action theory, estimates of internal consistency may not be applicable to the assessment of implicit motives in the first place. Atkinson essentially argued that internal consistency estimates are suitable for assessment tools, such as trait inventories, that tap declarative, self-related memories and beliefs. These frequently capitalize on a person's long-term memory of his or her past behavior, and on his or her short-term memory of answers given to earlier items on the same test; in conjunction with the compelling need to maintain a consistent self-concept, people are more likely to provide answers on subsequent items that are consistent with earlier ones (cf. Gazzaniga, 1985). But internal consistency measures are not suitable for the assessment of a motivational process, which by its very nature is not constant, but characterized by a dynamic waxing and waning of need states depending on environmental incentive cues and opportunities for need satisfaction. According to Atkinson, the PSE picks up such a motivational process and thereby indirectly taps the strength of an underlying motive disposition. The stronger the motive, the easier it is to arouse and show up in participants' stories and actual behavior. And then, by the very act of being expressed in imaginative thought (or actual behavior), the motive becomes consummated and drops for some time from the surface of participants' stories (or behavior). Atkinson, Bongort, and Price (1977) showed in computer simulations that individuals with a strong underlying motive disposition cycle more quickly back and forth between motive expression and satiation than individuals with a weak motive and predicted that, given equal PSE length, a motive therefore can be more validly assessed in the former than the latter. Reuman (1982) empirically confirmed this prediction of Atkinson's motivational theory by demonstrating that PSE motive scores with high variability, resulting in negative Cronbach alpha coefficients (!), have better predictive validity than motive scores with low variability and positive alphas.

A second source of inter-story variability of motive scores may be found in the dissimilarity of PSE picture cues. While Atkinson's dynamics-of-action model assumes a picture pool with similar, constant cue characteristics for a given motive, PSE picture sets used in research frequently consist of pictures showing a wide variety of situations (cf. Table 1). McClelland et al (1953) have speculated that a motive's strength is not only reflected in the intensity of individuals' responses to the same type of cue, but also in the extensity of the cues that elicit a motivational response, that is, how many different contexts and situations a person has learned to associate with need satisfaction. According to this view, individuals who, for instance, inject power imagery only into stories written in response to pictures showing couples, but not into stories about pictures showing competitive situations, may have learned to express their need for power only in close relationships, but not in competitive situations, whereas someone writing power-related stories about both types of pictures would have a more extensive power motive, because he or she responds to a greater variety of situations with power imagery. To our knowledge, there has been no systematic research on the extensity postulate. But if the postulate is correct, then lower internal consistency on the PSE would be due, at least in part, to a bandwidth-fidelity problem, according to which a PSE that samples motive imagery in response to highly dissimilar picture cues would by necessity have lower internal consistency, but broader validity, than a PSE with highly similar picture cues. We will return to this issue when we discuss picture selection.

Motive score reliability is high when examined from the perspective of coding agreements between two or more independent coders of the same PSE stories. Usually, high coding reliability is ensured from the outset, because most coding systems require the coder to train on practice stories prescored by an expert until they reach a criterion of .85 percentage agreement or better. A coder who has reached this criterion and follows further steps to ensure high coding reliability (as detailed below) typically also achieves good to excellent agreement with other coders who independently code the same PSE stories (cf. Pang & Schultheiss, in press; Smith, 1992b).

Motive score reliability is also satisfactory when viewed from the perspective of retest stability. To illustrate this point, we conducted a meta-analysis that included a comprehensive set of retest reliability studies, both published and unpublished, with the requirement that they employed motive scoring systems that were developed empirically in the McClelland-Atkinson tradition and used standard PSE administration conditions. The regression function we obtained for retest coefficients from these studies is 0.71 – 0.13 * base-10 log (retest interval in days; cf. Figure 1). According to this formula, the average stability coefficient is .71 if the retest interval is 1 day, .60 if the retest interval is 1 week, .52 is the retest interval is 1 month, .37 if the retest interval is 1 year, .29 if the retest interval is 5 years, and .25 if the retest interval is 10 years. Average retest stability did not differ across the motive domains power, achievement and affiliation/intimacy (collapsed into one category), and neither was motive stability differentially affected by motive domain across retest intervals (i.e., the slope was the same for all motive domains). Factors that also influence retest stability but we did not control for include test-retest instructions (e.g., allowing participants at the retest session to write stories that are similar to those they wrote at the first session increases stability; Lundy, 1985; Winter & Stewart, 1977), number and suitability of picture cues (longer tests and higher cue-strength pictures typically provide more stable or comprehensive estimates of an individual's true score; cf. Reumann, 1982), whether different pictures or the same pictures were used at both assessments (same pictures yield higher stability coefficients; cf. Lundy, 1985; Winter & Stewart, 1977), and the occurrence of motive-arousing or stressful life events during the test interval (critical life events can have a profound impact on motive scores and lower retest stability; cf. Smith, 1992b). Overall, our findings suggest that implicit motive scores are moderately stable over time and that stability decreases with increasing retest intervals at a rate similar to that observed for questionnaire trait measures (e.g., Schuerger, Zarrella, & Hotz, 1989).

III. Step-by-step guide for the assessment of implicit motives and the use of motive scores in analytic designs

In the following, we will provide detailed instructions for the assessment of implicit motives and the use of motive scores derived from the combined use of the PSE and implicit-motive coding systems developed in the McClelland-Atkinson research tradition. Although verbal cues are sometimes used instead of picture cues to elicit stories, we will focus on the PSE as the primary means of story collection, because it is the most widely used and best studied assessment tool for implicit motive measurement. In our recommendations, we will build both on the excellent chapter by Smith, Feld, and Franz (1992; see also the other chapters in Smith, 1992a, for recommendations for the assessment of specific motives) and our lab's own best practices, which we developed across dozens of studies and thousands of PSE stories (cf. Pang & Schultheiss, in press; Schultheiss & Brunstein, 2001). Our goal is to provide the uninitiated reader with robust guidelines for the collection of valid and reliable motive scores.

Which motives to assess?

The first question that needs to be answered in the planning of motive research is whether a single motive should be assessed – and if so, which – or whether several motives should be measured. The answers to these questions depend on the nature and focus of the research and will determine not only the choice of coding system(s), but also the number and kinds of pictures that will be used.

Therefore, let us start with brief descriptions of the most commonly used coding systems for *n* Achievement, *n* Affiliation, *n* Intimacy, and *n* Power (detailed versions of these coding systems can be found in Smith, 1992a; see also Schultheiss, 2001b; Winter, 1994). These coding systems have two common characteristics.

First, they were constructed either by examining the effects on story themes of experimental manipulation of the target motive, or by examining the differences in stories written by members of theoretically relevant, naturally occurring groups that differ in the strength of that motive. Second, each scoring system typically follows a general theoretical framework of the sequence for motivated behavior. The motivated behavioral sequence is often initiated when the person feels a need or a motive (*Need*). She may anticipate either successful attainment of this goal or she may become frustrated and anticipate non-attainment (*Goal Anticipation*). In order to accomplish her goal, she engages in instrumental activity that is either successful or not (*Instrumental Activity*). She may experience positive or negative affect in response to the goal-directed activity and its consequences (*Affect*).

Achievement-related coding systems. McClelland et al's (1953) scoring system for the achievement motive provided the blueprint for many other subsequent motive scoring systems. They introduced the technique of using experimental arousal to manipulate and examine the effects of a motive, and applied the general behavioral sequence outlined above to provide an organizing framework for their scoring system. The *n* Achievement scoring system also provided a guideline for making scoring decisions that would be adopted by many other authors—the scorer must first determine whether a story contains any reference to the motive goal before going on to score the motive-related subcategories. With respect to achievement, the scorer must determine whether the story contains reference to an achievement related goal in order to justify scoring achievement-related subcategories. For the *n* Achievement scoring system and other scoring systems described here, a scoring decision is made for the absence or presence (given a value of +1) of each scoring category and subcategory. Each category and subcategory can only be scored once per story. The motive score for each story is calculated by adding the category and subcategory scores for that picture and the motive score for each individual is computed by adding the motive scores for all pictures.

According to McClelland et al. (1953), stories are scored for achievement-related imagery if there is mention of *competition with a standard of excellence*. Having determined that a story contains achievement imagery, a coder then proceeds to score for seven possible subcategories: (a) stated need for achievement; (b) successful, doubtful, or unsuccessful instrumental activity; (c) positive and negative anticipatory goal states; (d) obstacles or blocks that can either come from either within the person or from elements in the world-at-large; (e) nurturant press, which refers to forces that aid the story character in the goal pursuit; (f) positive or negative affective states associated with goal attainment or nonattainment; and (g) achievement thema, which is a weighting category scored whenever achievement imagery is elaborated in such a way that it becomes the central plot of the story.

The latent distinction between a success-approaching and a failure-avoiding component of *n* Achievement inherent in McClelland et al.'s (1953) scoring system was later made explicit in Heckhausen's (1963) scoring system, which yields separate scores for hope of success and fear of failure. This scoring system, which was extensively validated with German samples (cf. Heckhausen, 1991), is now available in an English translation, which also includes extensive training materials (Schultheiss, 2001b).

Affiliation-related coding systems. There are two main approaches to measuring affiliative tendencies—the *n* Affiliation scoring system (Heyns, Veroff, & Atkinson, 1958), which was developed to assess concerns over establishing, maintaining, or restoring relationships, and the *n* Intimacy scoring system (McAdams, 1980), which assesses a preference for warm, close, and communicative social interactions. While *n* Affiliation is related to an active striving to attain and maintain social relationships, *n* Intimacy is focused on the quality of these relationships. In other words, intimacy involves a focus on the experience of "being" in a relationship, while affiliation emphasizes the act of "doing" or "getting" a relationship.

A story is scored for affiliation only if there is evidence of a concern over establishing, maintaining, or restoring a positive affective relationship with another person or persons. If a story qualifies meets this criterion, the scorer goes on to code for subcategories that are similar to those featured in McClelland et al.'s (1953) *n* Achievement scoring system. They include (a) stated need for affiliation; (b) instrumental activity; (c) anticipatory goal states; (d) obstacles or blocks; (e) affective states; and (f) Affiliation thema.

The scoring system for *n* Intimacy comprises two “prime tests” and eight subcategories. The scorer first determines whether a relationship produces positive affect or whether there is dialogue or any verbal or nonverbal exchange between (among) characters in a story. If either or both of these criteria are met, the scorer goes on to code for the subcategories: (a) psychological growth and coping; (b) commitment or concern by a story character for others; (c) time- or space-transcending quality of a relationship; (d) physical or figurative union of people who have been apart; (e) harmony; (f) surrender or acquiescence by a story character to uncontrollable outside forces that control interpersonal relationships; (g) escape to intimacy, which is when a story character escapes to a situation that affords greater interpersonal growth; (h) connections with the outside world, which is scored when at least one of the story characters experiences a metaphoric interaction with aspects of the nonhuman world (e.g., “communion with nature”) or a direct interaction that has a demonstrable effect on the character’s thought, behavior, or feelings (e.g., “the heat sapped his energy”).

Power-related coding systems. The original coding system for *n* Power (Veroff, 1957) was conceived of as a concern for controlling the means of influence. However, motive scores derived from this coding system were related to assertive behavior in some conditions (e.g., Veroff, 1957) but not in others (e.g., Berlew, 1961). These results were interpreted as indicating a *fear of weakness*, rather than a positive desire for power. In other words, people who received high scores on Veroff’s (1957) measure of *n* Power were interested in controlling the means of influence only to ensure that they would not become victims of others’ influence. In order to address the approach aspects of power, Winter (1973) constructed a revised scoring system for *n* Power which incorporates aspects of Veroff’s coding system and also of a more approach-oriented need for influence measure developed by Uleman (1972), but which goes beyond both in developing a broader view of the power motive as a concern for having impact on others. We therefore present a brief summary of Winter's revised coding system here and refer interested readers to Veroff (1957) and Ulemann (1972) for more information about earlier measures power motivation.

According to Winter (1973), power imagery is scored when a story character is concerned about having impact or influence over another person, persons, or the world-at-large. If power imagery is scored, a coder then goes on to check for the following subcategories: (a) prestige of actor; (b) stated need for power; (b) instrumental activity; (c) external blocks to power goal; (d) goal anticipation; (e) affective goal states; (f) effect, which refers to physical, mental, or emotional responses to the power attempts of a story character.

Winter's integrated scoring system. Winter's (1994) Manual for Scoring Motive Imagery in Running Text was constructed based on the systems developed by McAdams (1980) for *n* Intimacy, Heyns et al. (1958) for *n* Affiliation, McClelland et al. (1953) for *n* Achievement, and Winter (1973) for *n* Power. Each subscale of the running text system is an abbreviated version of the original motive system on which it is based. In other words, only the basic imagery of each motive is scored; the subcategories are not used. The running text system also combines *n* Affiliation and *n* Intimacy into one conjoint category because of the theoretical and empirical overlap between the two constructs.

The running text scoring system has been psychometrically validated (Winter, 1991) and demonstrates good predictive validity in a variety of applications ranging from studies of political leaders (Winter, 1980; Winter & Carlson, 1988) to the effects of motives-goal congruence on emotional well-being (Brunstein et al., 1998). Thus, it provides a viable alternative to the original coding systems, while substantially decreasing the amount of time and energy required of coders for training and research purposes. It was also developed to be used with other imaginative verbal material besides the PSE, and has been successfully used with, for instance, political speeches, literary documents, and spoken interviews (e.g., Winter, 1991; Winter & Carlson, 1988). Its versatility and ease of use make the Winter (1994) running text system a good choice for researchers who require a valid, albeit less differentiated, measure of intimacy/affiliation, achievement, and power motives.

What pictures to use?

We will now consider several issues in using pictures in PSE motive research, such as: the optimum number of picture cues to be presented; the influence of order of presentation on motive scores; and the selection of appropriate pictures for single- or multi-motive measurement.

Use a sufficient number of picture cues (at least 4 or 5). The two key issues surrounding number of pictures to include in a final PSE battery are validity and variance. Researchers should use batteries of eight pictures or fewer, because later stories tend to produce less valid scores than earlier pictures in a battery (Reitman & Atkinson, 1958). This decline in validity as test length increases is probably due to fatigue. The key is to use as few or as many pictures as the participant will cooperate in producing stories in one session, while not sacrificing validity of scores. Conversely, having too few pictures is not optimal either because variance of scores is affected as the number of pictures used in the PSE approaches zero. Logically, the degree of dispersion of scores within a population should increase proportionally with the number of items used in a test.

To illustrate this point, we have aggregated the motive scores from a dataset using a multi-motive PSE and encompassing five different studies (for details on this data set, cf. Pang & Schultheiss, in press), and expressed the distribution of motive scores as a function of number of pictures cues. Figure 2 shows the distribution of total motive imagery resulting from PSEs with as few as one and as many as six picture cues. As shown, the likelihood of detecting any motive imagery in a PSE protocol based on just one picture is less than 40%, and scores are Poisson-distributed, allowing the researcher only to derive a very crude dichotomy of the absence or presence of motive imagery, which is clearly too coarse and not adequate for most assessment and research purposes. Although the situation becomes slightly better in the case of two-, three-, and four-picture sets, the distribution of scores is nonetheless still considerably skewed to the left and, because too many individuals still show zero motive imagery, a normal distribution remains almost impossible to obtain, even if scores are subjected to square-root or log transformations. Motive scores start to approximate a normal distribution once a five-story picture set is used.

Thus, to the extent that one wants to, firstly, use motive scores in regression analyses, which require normally distributed predictor and criterion variables, and, secondly, make sufficiently fine-grained distinctions of motive strength among research participants, we would recommend that a PSE include at least five pictures. Note, however, that this recommendation is based on a data set using a multi-motive picture set (i.e., a PSE with pictures aiming at several different motives) and that the minimum number of pictures to use may be slightly lower (i.e., four) if a PSE is employed whose pictures all have a strong pull for one single motive.

Use pictures that have sufficiently high cue strength for the motive(s) of interest. Pictures can differ quite dramatically in their capacity to elicit motive imagery. Some pictures have a good "pull" for one kind of imagery (e.g., power), whereas others are suitable for assessing two or more motives, and yet others have very little pull for any motive. Moreover, some pictures suggest a motivational theme so explicitly that respondents will almost uniformly infuse their stories with a specific kind of motivational imagery and the picture therefore is not suitable to differentiate between individuals low and high in the motive in question. For instance, a picture that shows a boxer punching another boxer in the face is certain to elicit high amounts of power imagery from almost every respondent, because nearly everyone will write a story that describes the fight and the exchange of blows. On the other hand, a picture showing a boxer staring absent-mindedly into space may prompt some individuals to write about a fight ending in a knock-out (e.g., high-power individuals), but others may write about a boxer taking a break from practicing (e.g., low-power individuals) or about the boxer looking forward to spending the evening with his sweetheart (e.g., high-affiliation individuals). Thus, the former picture would be less suitable for discriminating high- from low-power individuals than the latter.

Clearly, therefore, what a researcher needs are picture cues that elicit a sufficient amount of motive imagery in respondents' stories, but that are sufficiently ambiguous or also provide some cues for other motives so that story content can also revolve around other themes than the motive in question. To facilitate the task of compiling suitable picture sets for the assessment of one or several motives, we have started to examine picture profiles, that is, to describe frequently used and new PSE pictures in terms of how much power, achievement and affiliation/intimacy imagery they elicit on average in respondents' stories (Pang & Schultheiss, in press; Schultheiss & Brunstein, 2001). We have compiled this information, which comes from US American (Pang & Schultheiss, in press) and German samples (Schultheiss & Brunstein, 2001), in Table 1 and complemented it with picture profiles from other studies that were conducted in our lab and that employed novel picture cues taken from newsmagazines. The stories these data are based on were all scored using Winter's (1994) integrated scoring system. Pictures that elicited at least one scoreable instance of imagery for a given motive in more than 50% of participants were considered high-pull picture for that motive. Thus, for instance, *women in laboratory* has high pull for both power and achievement imagery, whereas *ship captain* pulls primarily for power imagery and *couple by river* pulls primarily for affiliation imagery.

Usage of this information is straightforward: if you are interested in assessing one specific motive, select the 4 to 6 pictures that have the highest pull for that motive. For instance, a researcher interested in assessing only power could compile a set using the *ship captain*, *woman and man arguing*, *hooligan attack*, and *men on ship deck*. Based on the mean levels of power imagery given for each of these pictures in Table 1, the average sum score for power motive imagery will be close to 7, which is high enough to ensure a good spread of individual scores from the low end (i.e., close to 0) to the high end (i.e., 14 and higher) and thus to harvest a sufficient amount of variance representing interindividual differences in power motivation. More generally, for assessment of a single motive we recommend that at least 4 pictures are used that have high pull for that motive and for which a total score of 5 or more motive images can be expected on average. For the simultaneous assessment of two motives we recommend that at least 5 pictures are used that have high pull for either or both motives and for which total motive scores of 4 or more can be expected on average for each motive. And for the simultaneous assessment of all three major motives we recommend that at least 6 pictures be used, with each picture ideally having high pull for at least two motives, and with total motive scores being at 3 or higher for each motive. For this case, we (Pang & Schultheiss, in press) specifically recommended the use of the following set of pictures, which have been used and validated extensively in our lab: *couple by river*, *nightclub scene*, *women in laboratory*, *ship captain*, *trapeze artists*, and *boxer*. If picture sequence can not be randomized, we recommend presenting them in the sequence listed here.

Use pictures that are similar to the situation in which you assess your dependent variables. An issue that has received very little attention in the research literature is the relationship between the picture cues and the situational context in which dependent measures are being assessed. How similar to that situational context should the picture cues be to yield optimally valid motive scores? In lieu of systematic empirical investigations of this issue, we recommend that, to the extent that emotions, cognitions or behaviors are assessed as dependent variables in a specific situational context, the selected picture cues should be reasonably similar to, but not actual depictions of, the targeted situation. This recommendation is borne out by some observations in our lab. For instance, in a study on the effect of power motivation on persuasive communication (Schultheiss & Brunstein, 2001), the power motive score that best predicted participants' persuasiveness stemmed from the picture ship captain, which depicts a captain who seems to argue with another person, whereas the power motive score derived from the picture trapeze artists, which shows two trapeze artists in mid-flight, was a less valid predictor of persuasiveness, although when aggregated with the power motive scores from other pictures it contributed somewhat to the predictive validity of the overall power motive score. Conversely, power motive scores derived from cues that depicted individuals giving a presentation in front of a group of people held virtually no predictive validity in, and were therefore eliminated from, a set of studies in which the effects of power motivation on hormonal and behavioral responses to competition were assessed, whereas power motive scores stemming from pictures showing people in various aggressive or competitive situations had good predictive validity (Schultheiss, Wirth et al., 2005; Wirth et al, submitted).

If implicit motives are assessed in studies that do not focus on specific thought, feelings or behaviors in circumscribed situations but that aim at revealing the relationship of motives to more global measures (e.g., emotional well-being over time or life outcome variables; cf. Brunstein et al, 1998; Jenkins, 1997), it may be more useful to employ picture sets that represent a broad spectrum of everyday situations and thus capture the extensity of a person’s motives, that is, the range of contexts and situations in which the person’s motives become engaged and may ultimately influence the outcome measures of interest (cf. our previous discussion of motive intensity and extensity). The picture set described by Pang and Schultheiss (in press) represents a good example of a PSE with a broad range of validity.

Finally, keep in mind that researchers are not restricted to the traditionally used PSE pictures, like the ones described in Table 1 or contained in Smith (1992a). If a particular study requires the fine-tailoring of picture cues to a particular assessment situation or social context, new picture cues can be introduced by, for instance, using pictures from ads and reports in newspapers and magazines, as has been frequently done in the past (the picture *nightclub scene* represents is actually a beer ad from the 1960ies). However, we recommend that new pictures be pretested with a small sample of participants (20 to 30) to ensure that they have sufficient pull for the motive(s) in question. We also recommend that, if new pictures are used, they are complemented with a suitable selection of PSE cues whose validity is already established to minimize the risk of obtaining null findings due to possible low validity of the new cues.

Picture order has little influence on scores. Picture order has negligible effects on motive scores. As a general rule, picture profiles (i.e., how much power, affiliation, and achievement imagery they elicit) do not change differentially as a function of their position in the PSE set (Pang & Schultheiss, in press). Pang and Schultheiss (in press) found some evidence, however, that some pictures are more likely to elicit overall more motive imagery on some serial positions than others. The practical significance of such findings is probably negligible, as they were only obtained for two out of six pictures and the size of the effect was small. In general, picture order effects are of little concern if picture sequence can be randomized across participants by, for instance, administering the PSE on the computer. If picture order is fixed because, for instance, the PSE is administered as a printed leaflet, Smith et al. (1992) have recommended that pictures with low and high pull for a given motive should be alternated. If several high-pull pictures for the assessment of one motive only are administered, we recommend that similar pictures (e.g., pictures showing competitive situations) are not presented back to back, but interspersed with high-pull pictures with dissimilar content.

Test administration

The PSE can be administered individually, with picture presentation and timing of stories either being controlled by an experimenter or by a computer, or in small groups, with picture presentation and timing of stories typically being controlled by the experimenter. Each picture is presented for about 10 to 15 seconds, after which participants are told to write a story on a sheet of paper or type it directly into the computer. During the writing of the story, participants should no longer be able to see the picture so that their stories do not become confined to a mere description of the picture. After 4 min and 30 sec have elapsed since picture presentation onset, participants are told to finish their story, and when 5 min have elapsed since picture presentation onset, the next picture is presented. Thus, each picture presentation and story writing episode should last about 5 min. However, this is only a rough guideline. Inevitably, some participants will go overboard on a story and take more than the total allotted 5 min to write a story, and others will be done even before the 5 min are up. If assessment time is limited, but the researchers still would like to use a PSE with a large number of pictures, it is also possible to limit the total time allotted to each picture story to 4 min.

If the PSE is administered by an experimenter, we recommend that he or she exerts as little pressure as possible and moves on to the next picture presentation only when all participants are finished with a given story and make eye contact with the experimenter. If the PSE should not take longer than a specific amount of time in the overall data collection session, the experimenter can hold total PSE time constant by providing less time (or more time, in case participants finish early on a preceding picture) on subsequent pictures.

For computer administration of the PSE we recommend that the participant be reminded with a detectable, but low-key message (e.g., the words “Please finish your story and press a key to move on to the next picture” blinking on the screen, with a short beep repeating every 10 sec) once time is up on a given story. Moreover, to ensure that participants do not just skim through the PSE and produce insufficient story material in order to be finished early, we recommend that the software allow participants to move on to the next picture story only after some minimum time (e.g., 4 min) since picture onset has elapsed.

The overarching goal of PSE administration should be to exert *as little pressure as possible* on participants, because social pressure and demands can interact with participants’ motives in ways that renders the motive scores harvested from PSEs invalid (cf. Lundy, 1988). For this reason, we actually recommend computer administration of the PSE over administration by an experimenter, whose very presence and nonverbal signals may already impact on participants’ implicit motives (cf. Klinger, 1967). A viable, low-tech alternative is to use leaflets that contain all instructions and, in alternating order, picture cues and story writing sheets, and let participants work through the materials at their own pace, either in the lab or at home. Other measures that can decrease the pressure exerted on participants include: (a) not referring to the PSE as a “test” of any kind, (b) not exerting time pressure on participants by requiring them to "write quickly," or making ostensible use of a stopwatch or other time-tracking device (stopwatches should be used only covertly), (c) keeping the interaction between experimenter and participant relaxed and friendly but professional, and (d) administering the PSE at the beginning of a experiment, because other tests and tasks administered during data collection may have unintended, but systematic motive-arousing effects on participants (cf. Lundy, 1988).

Instructions. In our lab, we have worked with PSE instructions adapted from Lundy (1988), who obtained good motive score validity coefficients with them. These instructions are in turn based on earlier standard instructions given in Atkinson (1958) and Smith et al (1992). They read:

Picture Story Exercise

In the Picture Story Exercise, your task is to write a complete story about each of a series of [number of pictures] pictures - an imaginative story with a beginning, a middle, and an end. Try to portray who the people in each picture are, what they are feeling, thinking, and wishing for. Try to tell what led to the situation depicted in each picture and how everything will turn out in the end.

On your desk are six sheets of paper for you to write your stories on. They are labeled PSE 1 through PSE [number of pictures] in the upper right-hand corner. In the upper left-hand corner there are some guiding questions - these should be used only as guides to writing your story. You do NOT need to answer them specifically.

[For computer administration:]

Each picture will be presented for ten seconds. After it has disappeared, write whatever story comes to your mind. Don't worry about grammar, spelling, or punctuation - they are of no concern here. If you need more space, use the back of the sheet. You will have about five minutes for each story; the computer will then let you know when you have 20 sec left. If you take less than the entire five minutes, the computer will be ready to move on after four minutes.

[For administration by experimenter:]

I will show you each picture for ten seconds. After that, write whatever story comes to your mind. Don't worry about grammar, spelling, or punctuation - they are of no concern here. If you need more space, use the back of the sheet. You will have about five minutes for each story; I will let you know when you have 20 sec left and when it's time to move on to the next picture.

At this point, let us emphasize again that instructions should be conveyed as suggestions, not commandments. If an experimenter is presenting the instructions orally, he or she should be able to present their gist by heart rather than formally reading them off from a sheet.

In addition to the above instructions, a series of guiding questions are usually printed at the top of each writing page. These questions are adapted from Atkinson (1958) and read:

> What is happening? Who are the people?
> What happened before?
> What are the people thinking about and feeling? What do they want?
> What will happen next?

Although some investigators have spaced these questions apart evenly on the writing page in order to elicit specific responses to each question, we recommend against imposing this degree of structure to participants' responses, because it limits the free flow of the story narrative and may thus interfere with the expression of motivational impulses.

Sometimes participants are exposed to the same picture cues in two or more testing sessions. In these cases a sentence should be included in the instructions of the later session(s) that makes it explicit to participants that they are not expected to produce stories that are different from those they wrote before (Lundy, 1988; Winter & Stewart, 1977):

> You may remember seeing some of these pictures before. If you do, feel free to react to them as you did before or differently, depending on how you feel now. In other words, tell the story the picture makes you think of now, whether or not it is the same as you told last time.

Format of Data Collection. Usually, participants write their stories on designated sheets containing the guiding questions in the upper left-hand corner and the PSE story number in the upper right-hand corner to allow for easy matching of a given story to the eliciting picture cue later on. A relatively unexplored issue is the effect of having participants type their responses using the computer. Blankenship and Zoota (1998) compared power imagery in PSEs written by hand or on the computer, and found no differences in n Power in participants' stories between these conditions. This is good news because using the computer decreases the time required of human coders for transcription and word-counting tasks and makes data storage, organization, and sharing more convenient. Although stories typed in an untimed condition were longer than those that were typed in a timed condition, there were no significant differences in the motive scores. Thus, Blankenship and Zoota (1998) recommend that researchers use the untimed condition, because participants vary in their typing abilities.

Coding and preparation for coding

Coder Training. In order to move on from practice coding materials to actual protocols, scorers have to establish percentage agreement of 85% or better with materials prescored by an expert, as they are typically available with the training and practice materials of a coding system. Training materials for most scoring systems are reprinted in either Atkinson (1958) or Smith (1992a). If that is not the case, they can usually be requested from the author. Smith et al. (1992) have provided an excellent section on recommendations for training scorers to achieve adequate reliability with expert materials. A number of key points deserve to be highlighted: (a) read the coding manual several times before scoring practice materials until a sufficient level of reliability is obtained; (b) during training, review any errors before going on to the next set of practice stories; (c) novice scorers should undergo at least 12 hours of training with practice materials when learning a coding system for the first time; (d) construct, customize, and refer liberally to a "crib sheet" that contains the main points and definitions of each category and subcategory, important coding rules-of-thumb, and special coding considerations; (e) justify every coding decision by noting the category or subcategory that the image falls under. If a specific coding category cannot be identified, do not score the image no matter how relevant it seems. When in doubt, it is better to err on the side of caution and not score a statement for imagery. (f) Do not make physical markings on the original practice or research materials, in case there is a need to return to these materials later for re-scoring or re-training purposes. Sets of photocopies should be made for each coder and these may be marked up instead.

To these recommendations we add that if a novice coder does not achieve adequate reliability on the training materials at the first pass, he or she should score all training materials again, using a fresh set of unmarked copies, and as often as is necessary until the 85% criterion is reached. We do not see a problem in the fact that the coder will be better at the subsequent scoring passes simply because he or she remembers his or her own correct and incorrect scores on the first pass. Rather, this is exactly what a good coder should rely on: explicit memory for specific instances in which an image should or should not be scored. For the same reason, we also recommend that even a coder who has already passed the 85% hurdle and is starting to score stories from a research project keep the crib sheet or the original coding manual handy and refer to it *whenever even the slightest doubt arises* in the process of scoring. This will ensure that over time the coder will attain a full, explicit representation of the coding rules in all their complexity and be able to apply them with sufficient specificity and rigor.

An important statistic for coder training is an index of scoring reliability, which is the degree of scoring agreement between coders. Traditionally, there are two methods for calculating inter-scorer reliability: (a) Spearman's rank-order correlation between total motive scores and (b) percentage agreement, also known as index of concordance. Percentage agreement should be calculated for each motive using the following formula: [2*(agreements on presence of motive imagery in story)]/[number of times coder 1 scored motive imagery in story + number of times coder 2 scored motive imagery in story]. This formula is conservative because it does not take into account agreements on the absence of imagery.

Intraclass correlation coefficients (ICCs) are an increasingly endorsed way of assessing inter-rater reliability. The ICC is a “chance-corrected reliability coefficient suitable to continuous data and equivalent to kappa under appropriate conditions” (Meyer et al, 2002). The one-way random effects ICC is a form of analysis of variance that is calculated by the following formula: [MS (between raters) – MS (within raters)]/[MS (between raters) + MS (within raters)]. It is especially meaningful for calculating inter-rater reliability because it calculates correlations between observations that do not have an obvious order. In other words, in a study with two coders, the assignment of who is “rater 1” or “rater 2” is assumed to be random.

Coding. There are a couple of tasks and decisions involved in the preparation for coding. First, regardless of how long it has been since coders have mastered a coding system, they should re-acquaint themselves with it by re-reading the training manual before scoring any research materials. Doing so will improve subsequent coding accuracy and inter-rater reliability since individual scorers are likely to have developed idiosyncratic coding rules based on their previous coding experiences—a re-reading of the coding manual helps to alleviate the effects of this “scorer drift.” We recommend that after long breaks from scoring, coders re-establish reliability with practice materials before beginning to score research protocols.

Second, all materials for coding should be stripped of any identifying information such as participant biographical data and experiment condition. The only information that will appear on the protocols should be participant identification number and picture cue number. If the identification number discloses demographic information such as gender, race, or any other group identity, new identification numbers should be assigned. Careful blinding of protocols also prevents the “halo effect”. Occasionally, through scoring or being involved in other aspects of an experiment, a coder may form an impression of a participant that might influence her scoring of ambiguous responses by that participant. Smith et al. (1992) argue that one source of such a halo effect may simply be the coder’s knowledge of previous stories written by a participant, which may then influence the coder’s scoring of subsequent stories. They therefore recommend coding stories by picture cue as opposed to scoring them by participant. However, in our experience, careful application of the scoring rules usually prevents any carry-over from the scoring of previous stories to subsequent stories of a given participant. Moreover, scoring stories by cue instead of by participant may introduce error variance due to scorer drift across picture cues and also lead to more stereotypical scoring of stories due to similar story material. We therefore suggest that scoring by participant instead of picture cue is defensible, too.

Finally, one tedious but necessary task in addition to scoring for motive imagery is the determination of each story’s word count. Participants’ motive scores correlate with word count more than with the time participants have available to write each story or the time they actually take from the first word to the last, and word count is therefore the single most important variable to control for in participants’ overall motive scores.

One coder or two? A critical decision about coding concerns the number of coders to employ. We suggest that this decision should depend on coder experience. If a coder already has extensive coding experience (e.g., more than 1000 stories coded) with a given scoring system in actual research, *not* counting the training materials he or she completed, it is in our experience sufficient to have this one coder score stories from a new study, if the same scoring system and similar picture cues are used and the coder follows the recommendations we made in the section on *Coder Training*. In addition, we suggest that after coding of all materials is completed, the coder should review all of his or her scores by reading through the coded stories again and use this opportunity to eliminate scores that at second glance do not hold up to the coding rules and to detect images that may have been overlooked during coding and should be scored. This second reading is critical for maintaining high coding accuracy even if the coder is already experienced. Moreover, it may sometimes be necessary for the single coder to recode a subset of fresh, unmarked copies of 50 to 100 stories from the same sample to establish intra-rater reliability and demonstrate that coder drift has no substantial effect on the motive scores used in the data analysis.

If no experienced coder is available for coding, we recommend that two newly trained coders are used for independent coding of all stories. In our experience, inter-rater reliability can be substantially enhanced by first having both coders score independently 40 to 60 stories derived from the same picture cues as the actual sample's stories. These stories can either come from a subset (5 to 10 participants) of the actual sample or, better yet, from pilot participants whose data won't be used in testing the research hypotheses. After scoring of this initial story pool, the coders should meet, compare their stories, discuss all coding discrepancies, and try to develop further coding guidelines to deal with coding issues and problems that are idiosyncratic to the specific picture cues and sample they obtained their stories from. Of course, these specific guidelines should be consistent with the more general ones of the scoring manual employed by the coders. After that, each coder independently works on a full set of copies of the PSE stories collected from the total research sample.

Once all coding is finished, each story's total motive score can either be averaged between scorers for further analyses, or coders can get together for a final session in which coding discrepancies are discussed and resolved based on rigorous application of the rules contained in the scoring system. While the latter option is more labor intensive, it usually also provides more accurate and rationally derived final scores and has the added pedagogical benefit of further training for the coders. Whether the averaging or the discrepancy-resolution approach is chosen, inter-rater reliability should be determined based on each coder's original scores, before any changes are made to them due to averaging or coders' discussion of their scores. If intercoder reliability is determined using correlation or ICC coefficients, it should be done based on each participant's total score across all pictures, and not by picture.

A final comment on the effort invested into coding stories for motive imagery: On average, an experienced scorer needs 2 to 5 min to score one PSE story, not counting the time needed to determine word count (this task can be assigned to a research assistant). Thus, for a typical research sample of 80 participants who are administered a 6-picture PSE, an experienced coder will need between 16 and 40 hours to code all materials, plus some additional time to review the assigned scores. This is certainly a large time investment, but it is necessary to obtain valid motive scores, and the investment is very likely to pay off, as we have found again and again in our own research. We realize that, compared to the effort other sciences invest in their measurements, personality psychology often has an unfortunate tendency to sacrifice validity for economy. The PSE does not cater to this propensity, but neither is it excessively labor-intensive when compared to, for example, the assessment of hormones in psychoendocrinology or brain activation patterns in the cognitive and affective neurosciences.

Data entry and processing

Let us finish the step-by-step guide for implicit motive assessment by describing some data entry and processing strategies that we have found useful in our research. In the following, we will give recommendations about data entry, data aggregation and distribution issues, procedures for correcting for word count, and some basic approaches to data analysis using implicit motive scores.

Enter motive scores and word counts in as fine-grained a format as possible. At a minimum, motive and word count scores should be entered into a spreadsheet for each picture separately to enable the researcher to later examine the contribution of each picture to the PSEs criterion validity. In some cases, it may even be useful to enter subscores for each imagery category separately for each picture. This may be useful in cases in which, for instance, the researcher would like to compare the specific validity contributions of each coding category to the test's predictive power or separate the approach and avoidance components of a given motive, which are often represented as separate coding categories within the overall coding system (e.g., in McClelland et al's 1953 coding system for n Achievement, negative instrumental activity, negative goal anticipation, negative goal affect and the block categories could be used to assess the fear of failure aspect of achievement motivation; cf. Schultheiss & Brunstein, 2005, for further discussion).

As outlined in our previous discussion of reliability issues, researchers should not expect scores of the same motive derived from different picture cues to be correlated strongly. However, they can examine the differential validity of each picture's motive score by entering all pictures' motive scores, plus a word count variable representing the aggregated word count scores for each picture, into simultaneous regression equations with the dependent variables of interest. Not every picture score's regression weight will be significant, and sometimes even the signs of the picture scores may differ. However, before a specific picture's motive score is eliminated from a final total score, it should be firmly established that the picture's motive score differs from other pictures' scores because it (a) predicts a dependent variable with a different sign that is, moreover, (b) significant and (c) replicable for at least two critical criterion variables.

Examine score distributions after creating total motive and word count scores. After summing motive and word count scores across pictures, examine the total scores for each motive and the word count using histograms and scatterplots. If deviations from a normal distribution are detected or total scores feature outliers, use a square root or log transformation to correct the problem for the offending score. Keep in mind that outliers can substantially influence the results of later analyses that test the actual research hypothesis.

If motive scores and word counts correlate, correct motive score for word count. Calculate Pearson correlations for all total motive scores with the total word count score. If a given motive score does not significantly correlate with the total word count score, do not correct it for word count; use the raw total motive score in all further analyses. If a total motive score does significantly correlate with the word count score (usually in the positive direction and even up to .50), there are two strategies to remove the influence of participants' verbal fluency from the motive score.

The first option is to use regression analysis to residualize motive scores for word count and use the residual scores in all subsequent analyses. This option has the advantage that it makes the resulting motive score exactly 0-correlated with the word count score and thus completely removes the influence of verbal fluency from the motive measure. But there are also two disadvantages to this correction. First, because regression analysis is used, it depends on the entire sample's motive and word count scores to calculate predicted and residualized motive scores for each participant. Thus, a given residualized motive score's meaning is always sample-dependent. Second, the residualized scores lose the intuitive meaning of the raw score (an experienced researcher may easily classify a raw score of 5 as high, but may be stumped by a residual score of 0.89) and require subsequent transformations (e.g., into z scores) to make them more meaningful again. Note, however, that the meaning of a z score is, of course, entirely sample-dependent.

The second option is to correct motive score for word count by multiplying the total motive score with 1000 and dividing the result by the total word count for each participant. The resulting score can be readily interpreted as motive images per 1000 words. Due to the nature of the correction, the resulting score does not depend on the specific sample in which the data were collected and can easily be compared to average images-per-1000-word scores obtained in other samples. In other words, this transformation yields a readily interpretable metric that allows between-sample comparisons of motive scores. A major drawback of this correction is that the resulting scores are not necessarily 0-correlated with total word count. At worst, they can create serious biases and artifacts due to persisting significant overlap with the word count. Thus, the corrected motive score should always be checked for variance overlap with the total word count. The correlation will rarely be 0; however, we recommend that it should not exceed $|.15|$. If this criterion is not fulfilled, we recommend using the regression approach to word count correction.

Finally, check the word-count corrected motive score for distribution abnormalities using a histogram. If the corrected score is not normally distributed, but the raw total motive and word count scores were, use square root or log transformations to make the corrected score conform to a normal distribution. If the raw total motive or the raw word count score, or both, were not normally distributed to begin with, transform these first (or try a different transformation) so that the word-count corrected motive score will more closely approximate a normal distribution.

Use continuous word-count corrected motive scores in all statistical analyses. If total motive scores were carefully examined and corrected for either distribution violations, the influence of verbal fluency, or both, they will provide the researcher with a good quantitative measure of motive strength. Thus, word-count/distribution-corrected motive scores should be entered as quantitative variables into all correlation and regression analyses testing the research hypotheses and *not* be subjected to a median split or any other method of partitioning a quantitative measure into groups. This would only result in a rather dramatic reduction of test power, increase Type I error, and make it more difficult to test a research hypothesis. Usually, any kind of analysis that can be done with partitioned scores can also be done with quantitative scores when general regression methods are used. We will highlight this with a couple of case examples in the following and refer the reader to the excellent paper on the use of quantitative personality predictors in analytical designs by West, Aiken and Krull (1996) for a more systematic treatment of this topic.

IV. Case examples

In the following, we will showcase the use of PSE motive scores in a variety of research designs. Specifically, we will briefly highlight the use of the PSE in studies on motivational arousal, longitudinal changes in motive levels, and motive × situation interactions in between-subject designs.

Using the PSE to measure motivational arousal

Schultheiss, Wirth and Stanton (2004) were interested in studying the effects of affiliation motivation arousal on progesterone release. To do so, they had one group of participants view an affiliation-arousing movie, another one neutral control movie, and a third group a power-arousing movie to differentiate the specific effects of the affiliation-arousing movie from general effects due to the presentation of movies with social content. To verify that the different arousal conditions actually had the intended effect on participants' motivational state, they administered one 3-picture PSE before and one after the movie in a counterbalanced design, with one half of participants working on picture set A before and picture set B after the movie, and the other half working on the PSE in the reverse sequence. All PSE stories were scored for power and affiliation imagery using Winter's (1994) running text scoring system and motive scores were residualized for total word count on each set. Motive data were analyzed using a multiple regression analysis with post-movie motive scores as dependent variable, pre-movie motive scores as covariate, and experimental condition (control, affiliation arousal, power arousal) as dummy-coded predictor. The results verified that experimental conditions aroused implicit motives differentially: after the movie, affiliation-arousal participants had increased affiliation and decreased power motive scores, whereas power-arousal participants showed the reverse pattern, and control-group participants registered virtually no changes in their motive levels. Eventually, the affiliation-arousal group was also the only group that had elevated progesterone levels after the movie. In essence, Schultheiss et al (2004) used the assessment of motive changes on the PSE as a manipulation check, which allowed them to verify that the movie in the affiliation-arousal condition had the intended specific motivational effect. They therefore argued that the subsequent increase in progesterone observed in this condition reflected an affiliation-driven endocrine effect. Hormone assessments notwithstanding, their research design also represented a well-crafted study of the effects of motive arousal on changes in motivational content expressed in PSE stories and, in this aspect, is very similar to many motive arousal studies originally conducted to derive coding systems (e.g., Winter, 1973).

Assessing longitudinal changes in implicit motives

Schultheiss, Dargel, and Rohde (2003) examined the effects of menstrual cycle stage on longitudinal changes in implicit power and affiliation motivation. Their sample consisted of one group of normally cycling (NC) women, one group of women who were using oral contraceptives (OC), which prevent the hormonal changes typically associated with the various phases of the menstrual cycle from occurring, and one group of men. Three different sets of 4-picture PSEs were carefully compiled to ensure that each set featured picture cues that, while not identical to those in the other sets, depicted similar situations and had similar and balanced pull for the implicit needs for power and affiliation. These sets were administered at three phases of the menstrual cycle, corresponding to the follicular (T1), periovulatory (T2), and luteal phase (T3) of the female cycle (participants in the male group were scheduled based on a pseudo-cycle). To prevent a confoundation between picture set and assessment phase, Schultheiss et al (2003) systematically varied the sequence in which the sets A, B, and C were administered to participants across T1, T2, and T3, resulting in the six sequence permutations, ABC, ACB, BAC, BCA, CAB, and CBA. PSE stories were coded for power and affiliation imagery using Winter's (1994) integrated scoring system. All scores were converted to motive images per 1000 words to adjust for differences in word count. Although the three sets did not differ in the length of the stories written in response to them, set B elicited significantly more power imagery and significantly less affiliation imagery than the other two sets. To correct for these differences, participants' corrected motive scores were converted to *z* scores within each picture set. The influence of picture set on participants' motive scores in any given cycle phase was thus removed from the scores' variance. Repeated-measures regression analysis with motive (power, affiliation) and cycle phase (follicular, periovulatory, luteal) as within-subject factors and contraceptive use (OC, NC) revealed that women had higher levels of affiliation motivation around ovulation than in the follicular phase and lower power motivation in the periovulatory phase than in the luteal phase of the menstrual cycle, regardless of oral contraceptive use.

It is conceivable that the variance due to differences between picture sets may have obscured more fine-grained differences between groups and cycle phases. But Schultheiss et al (2003) decided against administering the same PSE three times to participants, because they were concerned that this procedure might introduce even more error variance due to participants getting bored with writing stories about the same pictures again and again. Thus, while this study illustrates that the PSE measure of implicit motives can also be used in longitudinal research, it also illustrates some of the problems associated with devising PSEs suitable for this task.

Examining motive × situation effects in an experimental study

A recent example of a study that examines the interplay between individuals' implicit motives and experimentally manipulated situations was provided by Schultheiss, Wirth, et al (2005). Two studies, the first with male, the second with female participants, dealt with the effects of power-relevant situational outcomes, namely, experimentally varied victory and defeat in a one-on-one competition based on an implicit sequence learning task. At the start of each data collection session, participants' implicit power motive was assessed with a 5-picture PSE. To optimize motive score validity for the prediction of physiological and behavioral changes associated with the competition situation created for this study, only picture cues were used that showed people in competitive or aggression-related contexts. The stories were coded for power imagery using Winter's (1973) revised power motive scoring system. Scores were corrected for word count by regression and the distribution of the z-transformed residuals checked for normality. In Study 2, a square-root transformation was used to make the z scores conform to a normal distribution. Before participants entered the contest, they listened to a tape-recorded guided imagery exercise that described the ensuing contest in rich contextual detail from the winner's perspective. The inclusion of this imagery exercise was based on earlier research indicating that mere verbal priming for and instructions about a potentially motive-relevant experimental task may not be sufficient to adequately engage participants' implicit motives in the typically language-heavy context of laboratory experiments (cf. our introduction and Schultheiss, 2001, for further discussion of this issue). Dependent variables were changes in salivary testosterone from before to after the contest and post-contest learning gains on the serial response task used during the competition. The analytical design was based on multiple regressions with power motive scores as a quantitative personality variable, contest outcome as dummy-coded experimental factor, and their multiplicative interaction term as predictors and post-contest testosterone (with pre-contest testosterone as a covariate) and implicit learning gains as dependent variables.

The power motive × contest outcome effect was significant for implicit learning gains in both studies. Among contest winners, power motives scores significantly and positively predicted implicit learning gains, whereas they were a significant negative predictor of learning among losers in both studies. In Study 1 (men), the power motive × contest outcome effect was also significant for testosterone 15 min post-contest, controlling for pre-contest testosterone levels. Among contest winners, higher power motives scores predicted, with marginal significance, higher testosterone increases; among contest lowers, higher power motives scores significantly predicted stronger testosterone decreases. Notably, like in an earlier study (Schultheiss & Rohde, 2002), post-contest testosterone increases were positively related to implicit learning gains, and mediation analyses revealed that the negative effect of power motivation on learning gains among losers was mediated by the effect of power motivation on testosterone changes. In Study 2 (women) power motivation had a strong main effect on post-contest testosterone changes, with higher motive scores predicting stronger hormone increases.

Ironically, although participants' self-reported mood was strongly influenced by contest outcome, with winners in both studies being much happier after the contest than losers, these changes in mood were neither predicted by participants' implicit power motivation, nor were they associated with hormone changes or learning gains. These findings underscore the point we were making previously, namely that non-declarative measures of motivation, such as implicit motive scores, implicit learning gains, or hormonal changes, are more closely associated with each other than with declarative measures, such as self-reported satisfaction or dissatisfaction with a situational outcome.

V. Conclusion and some future directions

We hope that the present chapter has made the procedures of measuring implicit motives with the PSE more transparent to potential users. To achieve this aim, we have relied on excellent earlier introductions to implicit motive assessment, which we recommend to anyone interested in motive assessment (e.g., Smith, 1992a; Smith et al., 1992), added recommendations based on our own experiences with PSE-based motive assessment (particularly with regard to the use of computers in the administration of the PSE), and compiled here for the first time crucial information on the motivational pull of traditionally used and more recent PSE picture cues. As we have emphasized throughout the chapter, valid assessment of implicit motives requires care, attention to detail, a thorough understanding of motivation in general and implicit motive dispositions in particular, and, yes, effort and time.

With regard to the last point, we should add that we do think it is possible to develop faster, and possibly also better, measures of implicit motives, as long as these reside in the non-declarative domain and assess motives by examining processes over which the individual has little conscious control and that tap some specific, crucial aspect of a given motive. One important step in this direction has recently been presented by Brunstein and Schmitt (2005), who adapted the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) to assess implicit achievement motivation. Consistent with the idea that implicit motives influence non-declarative, but not declarative indicators of motivation, these authors found that the IAT achievement motive measure was (a) uncorrelated with a self-report measure of achievement motivation based on the same stimulus material as the IAT measure and (b) predicted participants' response times on a vigilance task when performance feedback was provided, but not participants' self-reported satisfaction with their performance. Although the convergent validity of the IAT measure of achievement motivation with PSE measures of this motive still needs to be established, we think that Brunstein and Schmitt's findings are very promising and could mark the beginning of a new era of implicit motive assessment.

Still, even if latency-based or other procedural measures of implicit motives will be developed more intensively in the future, we are certain that there will always be a place for PSE-type motive measures for two important reasons. First, the content-coding systems developed for the PSE can also be applied to many other forms of verbal material and therefore enable the researcher to assess at a distance the personality of individuals who are either dead or in other ways inaccessible (cf. Winter, 1991). This simply can not be done with computerized, procedural motive measures, which always require the respondent to be present for the assessment. Second, those studies that use PSE-type methods to collect verbal material represent an immensely valuable resource, because stories can always be recoded once other or better content-coding measures are developed and become available to the researcher. Many old longitudinal studies that have employed the TAT for clinical or psychodynamic assessment purposes today represent a gold mine for motive researchers who code the TAT stories for implicit motive imagery to examine the impact of motives on life trajectories. Had the original authors of these studies used tests with a fixed response format instead of the open-ended collection of picture stories, later researchers would always be stuck with the necessarily more limited conceptual framework of the study authors and the data set would therefore be of much more limited value.

For these reasons, we believe that for a long time to come, the development and refinement of content-coding methods for motive assessment, including the PSE technique, will remain an important task for researchers who are interested in the nonconscious sources of motivation and personality. Some important ways in which the content-coding approach to the assessment of motivation could be furthered include the development of scoring systems for neglected, but fundamental motivational needs, such as sex or curiosity, the refinement of existing motive scoring systems, and the development of coding systems that, based on experimental arousal studies, systematically differentiate between approach and avoidance components of each motive (cf. Schultheiss, in press; Schultheiss & Brunstein, 2005). Ultimately, such efforts towards better assessment of implicit motives are likely to pay off in the form of more meaningful findings, more sophisticated theories of implicit motivational processes, and a more prominent role of the implicit motive construct for personality psychology and neighboring disciplines, such as clinical psychology and affective neuroscience.

References

(* denotes studies that were included in the meta-analysis)

Atkinson, J. W. (Ed.). (1958). Motives in fantasy, action, and society. Princeton, NJ: Van Nostrand.

Atkinson, J. W. (1981). Studying personality in the context of an advanced motivational psychology. *American Psychologist, 36*, 117-128.

Atkinson, J. W., & Birch, D. (1970). *The dynamics of action*. New York: Wiley.

Atkinson, J. W., Bongort, K., & Price, L. H. (1977). Explorations using computer simulation to comprehend thematic apperceptive measurement of motivation. *Motivation and Emotion, 1*, 1-27.

Berlew, D. E. (1961). Interpersonal sensitivity and motive strength. Journal of Abnormal and Social Psychology, 63, 390-394.

Biernat, M. (1989). Motives and values to achieve: Different constructs with different effects. *Journal of Personality, 57*, 69-95.

* Birney, R. C. (1959). The reliability of the achievement motive. *Journal of Abnormal and Social Psychology, 58,* 266-267.

* Birney, R. C., Burdick, H., & Teevan, R. C. (1969). *Fear of failure*. New York: Van Nostrand-Reinhold Company.

Blankenship, V., & Zoota, A. L. (1998). Comparing power imagery in TATs written by hand or on the computer. Behavior Research Methods, Instruments, and Computers, 30(3), 441-448.

Brunstein, J. C., & Hoyer, S. (2002). Implizites und explizites Leistungsstreben: Befunde zur Unabhängigkeit zweier Motivationssysteme [Implicit versus explicit achievement strivings:

Empirical evidence of the independence of two motivational systems]. *Zeitschrift für Pädagogische Psychologie, 16*, 51-62.

Brunstein, J. C., & Maier, G. W. (in press). Implicit and self-attributed motives to achieve: Two separate but interacting needs. *Journal of Personality and Social Psychology*.

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.

Brunstein, J. C., Schultheiss, O. C., & Grassmann, R. (1998). Personal goals and emotional well-being: The moderating role of motive dispositions. Journal of Personality and Social Psychology, 75, 494-508.

* deCharms, R. (1976). *Enhancing motivation in the classroom*. New York: Irvington.

deCharms, R., Morrison, H. W., Reitman, W., & McClelland, D. C. (1955). Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation* (pp. 414-423). New York: Appleton-Century-Crofts.

Entwisle, D. R. (1972). To dispel fantasies about fantasy-based measures of achievement motivation. *Psychological Bulletin, 77*, 377-391.

Fineman, S. (1977). The achievement motive construct and its measurement: Where are we now? *British Journal of Psychology, 68*, 1-22.

* Franz, C. E. (1994). Does thought content change as individuals age? A longitudinal study of midlife adults. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality change?* (pp. 227-250). Washington, DC: American Psychological Association.

Gazzaniga, M. S. (1985). *The social brain: Discovering the networks of the mind*. New York: Basic Books.

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

* Heckhausen, H. (1963). *Hoffnung und Furcht in der Leistungsmotivation [Hope and fear components of achievement motivation]*. Meisenheim am Glan: Anton Hain.

Heckhausen, H. (1991). *Motivation and action*. Berlin, Germany: Springer.

Heyns, R. W., Veroff, J., & Atkinson, J. W. (1958). A scoring manual for the affiliation motive. In J. W. Atkinson (Ed.), Motives in fantasy, action, and socety (pp. 205-218). Princeton, NJ: Van Nostrand.

Hofer, J., & Chasiotis, A. (2003). Congruence of life goals and implicit motives as predictors of life satisfaction: Cross-cultural implications of a study of Zambian male adolescents. *Motivation and Emotion, 27*, 251-272.

Jackson, D. N. (1984). *Personality Research Form* (3 ed.). Port Huron, MI: Sigma Assessment Systems, Inc.

* Jenkins, S. R. (1987). Need for achievement and women's careers over 14 years: Evidence for occupational structure effects. *Journal of Personality and Social Psychology, 53*, 922-932.

* Jenkins, S. R. (1994). Need for power and women's careers over 14 years: Structural power, job satisfaction, and motive change. *Journal of Personality and Social Psychology, 66*, 155-165.

Kagan, J. (2002). *Surprise, uncertainty, and mental structures*. Cambridge, MA: Harvard University Press.

* Kagan, J., & Moss, H. A. (1959). Stability and validity of achievement fantasy. *Journal of Abnormal and Social Psychology, 58*, 357-364.

Klinger, E. (1967). Modeling effects on achievement imagery. *Journal of Personality and Social Psychology, 7*, 49-62.

* Kraiger, K., Hakel, M. D., & Cornelius, E. T. (1984). Exploring fantasies of TAT reliability. *Journal of Personality Assessment, 48*, 365-370.

Langner, C. A., & Winter, D. G. (2001). The motivational basis of concessions and compromise: Archival and laboratory studies. *Journal of Personality and Social Psychology, 81*, 711-727.

* Lundy, A. (1985). The reliability of the Thematic Apperception Test. Journal of Personality Assessment, 49, 141-145.

Lundy, A. (1988). Instructional set and Thematic Apperception Test validity. Journal of Personality Assessment, 52(2), 309-320.

McAdams, D. P. (1980). A thematic coding system for the intimacy motive. Journal of Research in Personality, 14, 413-432.

McClelland, D. C. (1987). Human Motivation. NY: Cambridge University Press.

* McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). The achievement motive. NY: Appleton-Century-Crofts.

* McClelland, D. C., Davis, W. N., Kalin, R., & Wanner, E. (1972). *The drinking man*. New York: Free Press.

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

Meyer, G. J., Hilsenroth, M. J., Baxter, D., Exner, J. E., Jr., Fowler, J. C., Piers, C. C., et al. (2002). An examination of interrater reliability for scoring the Rorschach Comprehensive System in eight data sets. *Journal of Personality Assessment, 78*(2), 219-274.

* Morgan, H. H. (1953). Measuring achievement motivation with "picture interpretations". *Journal of Consulting Psychology, 17*, 289-292.

Murray, H. A. (1943). Thematic Apperception Test manual. Cambridge, MA: Harvard University Press.

* Niitamo, P. (1999). *"Surface" and "depth" in human personality: Relations between explicit and implicit motives* (People and work No. 27). Helsinki: Finnish Institute of Occupational Health.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Pang, J. S., & Schultheiss, O. C. (in press). Assessing implicit motives in U. S. college students: Effects of picture type and position, gender and ethnicity, and cross-cultural comparisons. Journal of Personality Assessment.

Reitman, W. R., & Atkinson, J. W. (1958). Some methodological problems in the use of thematic apperceptive measures of human motives. In J. W. Atkinson (Ed.), Motives in fantasy, action, and society (pp. 664-683). Princeton, NJ: Van Nostrand.

Reuman, D. A. (1982). Ipsative behavioral variability and the quality of thematic apperceptive measurement of the achievement motive. *Journal of Personality and Social Psychology, 43*(5), 1098-1110.

Schuerger, J. M., Zarrella, K. L., & Hotz, A. S. (1989). Factors that influence the temporal stability of personality by questionnaire. *Journal of Personality and Social Psychology, 56*, 777-783.

Schultheiss, O. C. (2001a). An information processing account of implicit motive arousal. In M. L. Maehr & P. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 12: New directions in measures and methods, pp. 1-41). Greenwich, CT: JAI Press.

Schultheiss, O. C. (2001b). *Manual for the assessment of hope of success and fear of failure (English translation of Heckhausen's need Achievement measure)*. Department of Psychology, University of Michigan, Ann Arbor: Unpublished manuscript.

Schultheiss, O. C. (in press). A biobehavioral model of implicit power motivation arousal, reward and frustration. In E. Harmon-Jones & P. Winkielman (Eds.), *Fundamentals of Social Neuroscience*. New York: Guilford.

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., & Brunstein, J. C. (2001). Assessing implicit motives with a research version of the TAT: Picture profiles, gender differences, and relations to other personality measures. Journal of Personality Assessment, 77(1), 71-86.

Schultheiss, O. C., & Brunstein, J. C. (2002). Inhibited power motivation and persuasive communication: A lens model analysis. *Journal of Personality, 70*, 553-582.

Schultheiss, O. C., & Brunstein, J. C. (in press). An implicit motive approach to competence. In A. J. Elliot & C. S. Dweck (Eds.), Handbook of competence and motivation. NY: Guilford.

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), 234-41.

Schultheiss, O. C., Dargel, A., & Rohde, 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., & Hale, J. A. (2005). *Implicit motives modulate attentional orienting to perceived facial expressions of emotion*. Manuscript submitted for publication.

Schultheiss, O. C., Pang, J. S., Torges, C. M., Wirth, M. M., & Treynor, W. (2005). Perceived facial expressions of emotion as motivational incentives: Evidence from a differential implicit learning paradigm. *Emotion, 5*(1), 41-54.

Schultheiss, O. C., & Rohde, W. (2002). Implicit power motivation predicts men's testosterone changes and implicit learning in a contest situation. Hormones and Behavior, 41, 195-202.

Schultheiss, O. C., Wirth, M. M., & Stanton, S. (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.

* Siegel, P. & Weinberger, J. (1998). Capturing the "mommy and I are one" merger fantasy: The oneness motive. In R. F. Bornstein & J. M. Masling (Eds.), *Empirical perspectives*

on the psychoanalytic unconscious (pp. 71-97). Washington, DC: American Psychological Association.

* Skolnick, A. (1966). Stability and interrelations of thematic test imagery over 20 years. *Child Development, 37*(2), 389-396.

Smith, C.P. (1992a). Motivation and Personality: Handbook of thematic content analysis. NY: Cambridge University Press.

* Smith, C. P. (1992b). Reliability issues. In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis* (pp. 126-139). New York: Cambridge University Press.

Smith, C. P., Feld, S. C., & Franz, C. E. (1992). Methodological considerations: Steps in research employing content analysis systems. In C. P. Smith (Ed.), Motivation and Personality: Handbook of thematic content analysis (pp. 515-536). NY: Cambridge University Press.

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin, 112*, 140-154.

* Uleman, J. S. (1972). The need for influence: Development and validation of a measure, in comparison with need for power. *Genetic Psychology Monographs, 85*, 157-214.

Veroff, J. (1957). Development and validation of a projective measure of power motivation. *Journal of Abnormal and Social Psychology, 54*, 1-8.

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality, 64*, 1-48.

* Winter, D. G. (1973). The power motive. NY: Free Press.

Winter, D. G. (1980). Measuring the motives of southern African political leaders at a distance. Political Psychology, 2(2), 75-85.

* Winter, D. G. (1991). Measuring personality at a distance: Development of an integrated system for scoring motives in running text. In D. J. Ozer, J. M. Healy & A. J. Stewart (Eds.), *Perspectives in personality* (Vol. 3, pp. 59-89). London: Jessica Kingsley.

Winter, D. G. (1994). Manual for scoring motive imagery in running text. Unpublished instrument, University of Michigan, Ann Arbor.

Winter, D. G. (1999). Linking personality and "scientific" psychology: The development of empirically derived Thematic Apperception Test measures. In L. Gieser & M. I. Stein (Eds.), *Evocative images: The Thematic Apperception Test and the art of projection* (pp. 107-124). Washington, DC: American Psychological Association.

Winter, D. G., & Carlson, L. (1988). Using motive scores in the psychobiographical study of an individual: The case of Richard Nixon. Journal of Personality, 56, 75-103.

* Winter, D. G., & Stewart, A. J. (1977). Power motive reliability as a function of test-retest instructions. Journal of Consulting and Clinical Psychology, 45, 436-440.

Wirth, M. M., Welsh, K. M., & Schultheiss, O. C. (submitted for publication). Salivary cortisol changes in humans after winning or losing a dominance contest depend on implicit power motivation.

Woike, B. A. (1995). Most-memorable experiences: Evidence for a link between implicit and explicit motives and social cognitive processes in everyday life. *Journal of Personality and Social Psychology, 68*, 1081-1091.

Zurbriggen, E. L. (2000). Social motives and cognitive power-sex associations: Predictors of aggressive sexual behavior. *Journal of Personality and Social Psychology, 78*, 559-581.

Table 1

Raw Motive Scores and Word Counts Across Picture Cues

	N Power		*N* Achievement		*N* Affiliation		Words	
Picture	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Women in Laboratory								
U.S.[a]	0.77	0.85	1.08	0.93	0.19	0.50	93.38	26.09
German[b]	0.80	0.84	0.66	0.77	0.19	0.48	82.85	21.15
Ship Captain								
U.S.[a]	1.01	0.88	0.14	0.47	0.21	0.53	98.24	27.46
German[b]	1.16	0.92	0.11	0.37	0.20	0.53	85.68	21.08
Couple by River								
U.S.[a]	0.23	0.54	0.00	0.21	2.06	1.07	99.42	25.73
German[b]	0.43	0.72	0.03	0.17	1.84	1.05	91.53	22.01
Trapeze Artists								
U.S.[a]	0.70	0.79	0.76	0.83	0.49	0.80	97.00	26.96
German[b]	0.79	0.85	0.78	0.84	0.43	0.71	86.04	21.64
Nightclub Scene								
U.S.[a]	0.75	0.82	0.01	0.30	1.32	1.10	99.53	28.09
German[b]	0.86	0.83	0.09	0.31	1.29	1.08	89.44	22.77
Boxer[c]								
U.S.[a]	0.79	0.90	1.14	1.06	0.17	0.51	97.60	29.09
Architect at Desk[c]								
German[b]	0.22	0.46	0.29	0.55	1.16	0.84	90.75	22.73

Table 1 (continued)

	N Power		*N* Achievement		*N* Affiliation		Words	
Picture	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Bicycle Race								
U.S.[c]	0.95	0.95	1.65	1.08	0.16	0.41	109.61	27.27
Woman and Man Arguing								
U.S.[c]	1.70	0.94	0.31	0.66	0.18	0.53	102.46	27.23
Hooligan Attack								
U.S.[c]	2.26	0.95	0.01	0.14	0.15	0.43	109.36	27.86
Lacrosse Duel								
U.S.[c]	0.69	0.84	1.22	0.96	0.16	0.50	102.77	27.08
Men on Ship Deck								
U.S.[d]	1.82	1.26	0.48	0.63	0.20	0.51	110.52	24.71
German[e]	1.14	0.95	0.47	0.73	0.29	0.63	77.49	18.61
Soldier								
German[e]	0.94	1.01	0.35	0.64	0.17	0.36	81.52	21.81
Soccer Duel								
German[e]	0.74	0.77	1.79	0.81	0.07	0.27	80.47	18.28
Couple Sitting Opposite a Woman								
U.S.[f]	1.05	0.98	0.40	0.79	0.73	0.90	104.45	23.84
Girlfriends in Café With Male Approaching								
U.S.[f]	0.53	0.65	0.38	0.76	1.62	1.43	105.45	22.42

Table 1 (continued)

Note. Underlined motive scores indicate that more than 50% of participants have responded with at least one instance of codable motive imagery to the picture cue.

[a] From: Pang & Schultheiss (in press); N = 323

[b] From: Schultheiss & Brunstein (2001); N = 428

[c] From: Wirth, Welsh, & Schultheiss, Study 2 (submitted); N = 109

[d] From: Schultheiss, Campbell, & McClelland (1999); N = 42 (male participants only)

[e] From: Schultheiss & Rohde (2002); N = 66 (male participants only)

[f] From: Schultheiss, Wirth & Stanton (2004); N = 60

Figure Captions

Figure 1. Regression line (plus 95% confidence interval) for motive stability coefficients as a function of test-retest time interval (log-transformed).

Figure 2. Distribution of motive images as a function of number of pictures in a picture set.

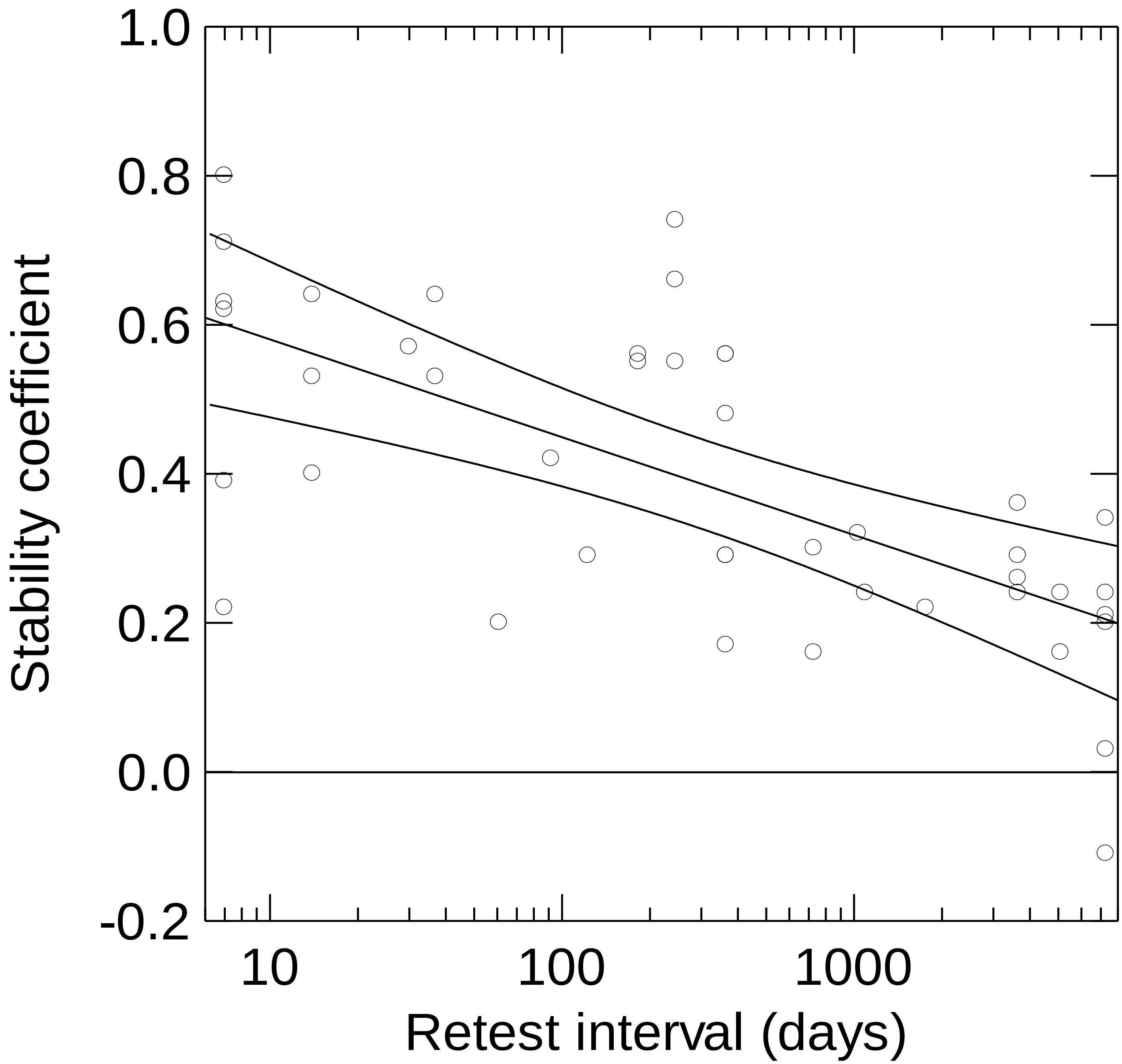
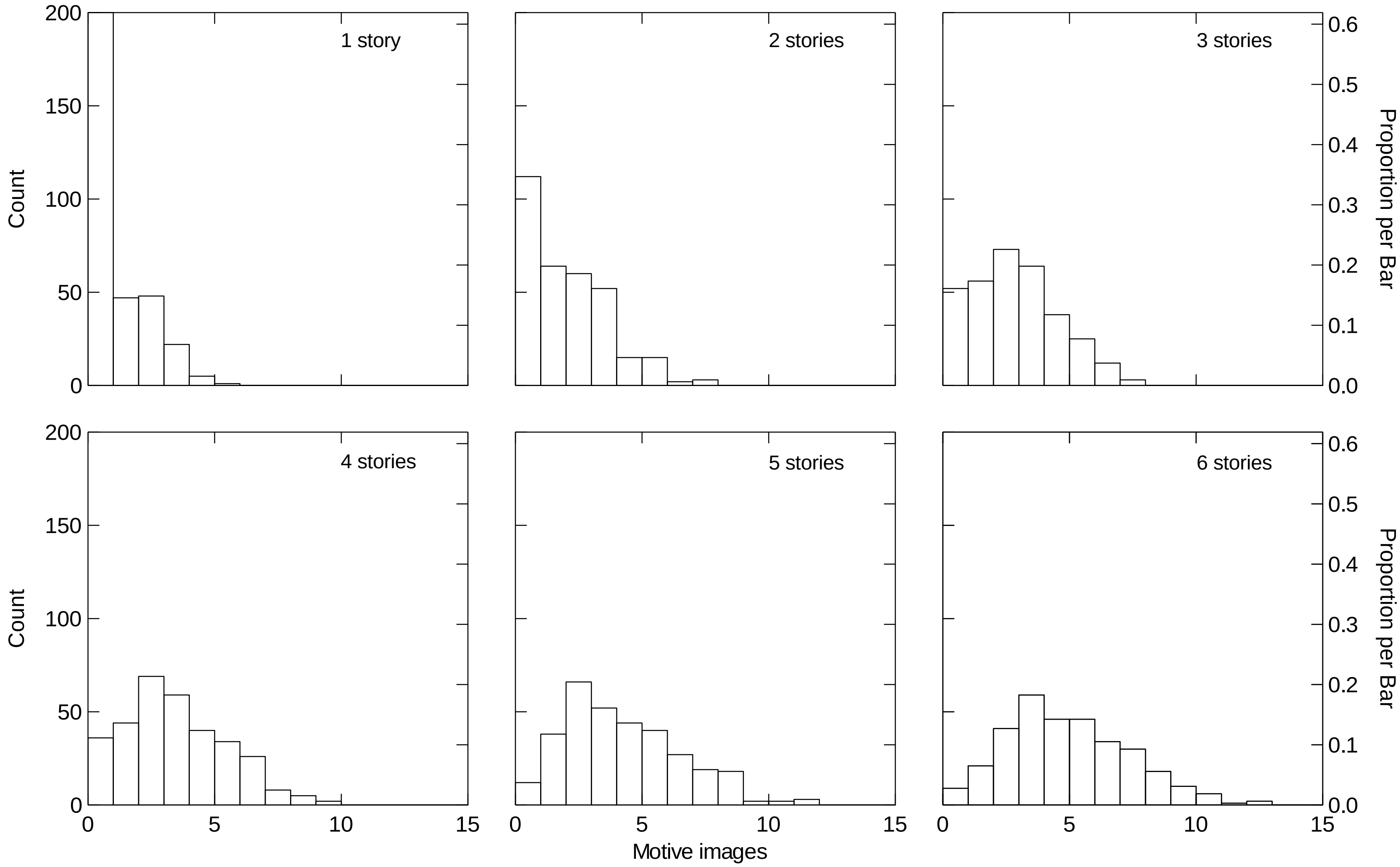

Stability coefficient
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
10
100
1000
Retest interval (days)

Figure 1.



Count
Proportion per Bar
1 story
2 stories
3 stories
4 stories
5 stories
6 stories
200
150
100
50
0
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
Motive images

Figure 2.