Self-Control in the General Theory of Crime:: Theoretical Implications of a Measurement Problem
Bernd Marcus
Theoretical Criminology 2004; 8: 33
DOI: 10.1177/1362480604039740

The online version of this article can be found at:
http://tcr.sagepub.com/cgi/content/abstract/8/1/33
Self-control in the General Theory of Crime:  
*Theoretical implications of a measurement problem*

**BERND MARCUS**  
Chemnitz University of Technology, Germany

**Abstract**

Numerous studies have attempted to test Gottfredson and Hirschi’s General Theory of Crime. The present article outlines the view that virtually every empirical test of the theory is based on serious misinterpretations of its core construct, self-control. A re-interpretation of self-control is proposed and seven requirements for its construct-valid measurement are specified. A review of self-control measures used in previous research shows that these requirements are more often violated than met. As a consequence, the empirical status of self-control theory is held to be still largely unknown, despite all apparent evidence.

**Key Words**

crime theory • personality assessment • self-control

In the few years after Gottfredson and Hirschi published their *General Theory of Crime* (1990, hereafter referred to as GTC), the book attracted an enormous volume of scholarly articles dealing with its empirical value for understanding criminal behavior, its theoretical and practical implications and often with some fundamental controversies around its core
conceptions. Due to this attention, however it may be motivated, the authors were ranked at top positions among the most-cited criminologists in recent years (Cohn and Farrington, 1999). Pratt and Cullen (2000) found 28 published empirical tests of GTC’s explanatory power for their meta-analysis, not counting numerous dissertations and other unpublished papers. Subsequent to this quantitative review, the stream of empirical research has continued to pour into scientific journals (e.g. LaGrange and Silverman, 1999; Tibbetts and Myers, 1999; DeLisi, 2001; Gibson and Wright, 2001; Vazsonyi et al., 2001; Szokjy and Geis, 2002; Turner and Piquero, 2002; Unnever and Cornell, 2003). It may appear, then, as if the GTC should have reached a state of settlement within criminological theorizing, allowing for a well-informed evaluation of its empirical status.

In this article, I try to show that such a conclusion may be premature. That is, I will not engage myself in discussions as to whether the GTC is simple (parsimonious) or simplistic, how it fares relative to alternative approaches, what are the political implications or ideological underpinnings of the theory, or other questions that may have contributed to the attractiveness of the GTC for both its proponents and critics. I will rather argue that all these discussions essentially lack both conceptual and empirical grounds up to this point and are therefore almost purely speculative. Because Gottfredson and Hirschi propose just one personality trait, self-control, as the core explanation for crime and explicitly reject most alternative theoretical approaches, it is essential to be as precise as possible about the meaning of that trait in discussing the GTC theoretically, and it is equally important to measure this trait adequately in testing the theory empirically. Both objectives require a rigorous translation of the theoretical concept into an operationally defined personality construct and further transmission into a measure that taps into the trait and not into something else. Upon finishing this process, we will be provided with the means to discuss and empirically test the fundamental controversies around the GTC. The main objective of the present article is to demonstrate that we are currently far from reaching this point and to outline what can and needs to be done to get there. This is not possible without the treatment of questions which are ‘technical’ in the sense of making the methodological implications of Gottfredson and Hirschi’s concept of self-control explicit. Readers of this theoretical journal who may be used to keeping questions of theory separate from methodological issues may see this article as a case study on the inextricability of theory and measurement in the social sciences. What I will try to show is that, in the case of the GTC, imprecise construct definition led to inadequate measurement and finally into essentially meaningless discussions around its theoretical value.

More specifically, it is assumed here that there is a logical order in the sequence of problems that any meaningful test of the GTC has to solve: (1) defining the term ‘self-control’ theoretically; (2) translating this definition into a measurement concept; (3) establishing that measure’s construct validity; and (4) relating self-control to outcome variables. Any step in this
sequence can provide an important independent contribution to the field if—and only if—the preceding steps are adequately addressed. Otherwise, any results will not be interpretable as tests of the theory. As I will try to demonstrate, the approaches currently predominating the literature on the GTC are seriously flawed at the conceptual stages (1), or (2), or both. When the independent variable is ill-defined, the numerous attempts to empirically test hypotheses on the relationships between self-control and dependent variables (step 4) cannot be relied on. Consequently, this article is concerned with steps (1) and (2) in order to provide a more promising basis for the empirical steps (3) and (4), respectively.

The article is organized into two major sections that roughly correspond to the objectives mentioned above. I begin with a critical review of present approaches to the measurement of self-control. In this section, both theoretical and methodological implications of Gottfredson and Hirschi's development of the GTC are discussed that are held to be logically inconsistent with some conceptual and psychometric characteristics of the more prominent self-control measures to date (step 1). I continue with the derivation of seven requirements for an adequate self-control measure, based on my interpretation of the GTC (step 2). The article concludes with a synopsis of the degree to which these requirements are met in previous research and a discussion of the implications for evaluating the current status of the GTC.

A critical review of existing approaches to the measurement of self-control

‘Self-control’ and related traits have a long and complex tradition in personality research as well as criminology (see Pulkinnen, 1986, for a review). For our purpose, however, more important than this label and the definitions other scholars provided for it is the way Gottfredson and Hirschi derived their particular understanding of the variable underlying criminal and analogous acts. The authors do not appear to place much value on previous psychological concepts of self-control but rather start their theory development by subscribing to the Benthamian conception of human nature as being ruled by pleasure and pain. They later outline their conception for the nature of crime and finally arrive at an explanation for crime in which they try to combine these two concepts. It states that criminals pursue the goals of seeking pleasure and avoiding pain like anyone else but that they lack a quality which makes most other people aware of the long-term pains inherent in the nature of crime that outweigh the short-term pleasures. Gottfredson and Hirschi label this construct self-control and emphasize its status as the core concept within their theory. Any comprehensive test of the theory will have to employ some means to operationalize self-control. If there is substantial incongruence between the theoretical derivation and the operationalization, it must be concluded that
The theory was not adequately tested. Obviously, a theory can neither be supported nor rejected without testing it in an appropriate manner. Thus, the first requirement in theory testing always is to show that the instruments used measure what they purport to measure. What makes up the construct of self-control has been subject to almost any published research report as well as theoretical commentaries, including some by the original authors (see particularly Hirschi and Gottfredson, 1993, 1994, 1995). Upon inspection, the original conception differs considerably from the way self-control was actually measured in most empirical studies.

The field of interest here, empirical tests of Gottfredson and Hirschi’s theory involving self-control as independent variable, can be roughly divided into two categories or ‘generations’ of studies. The first generation employed mostly secondary analyses of data sets collected before the GTC was published (Benson and Moore, 1992; Brownfield and Sorensen, 1993; Keane et al., 1993; Larragoite, 1994; Polakowski, 1994; Strand and Garr, 1994; Uihlein, 1994; Sorenson and Brownfield, 1995; more recent examples are Avakame, 1998; Paternoster and Brame, 1998), covering a quite diverse set of indicators for self-control (see Gibbs and Giever, 1995, for a table summarizing early self-control measures). I will restrict my discussion to the second generation of GTC studies, which employed questionnaire measures of self-control specifically designed to operationalize Gottfredson and Hirschi’s construct. Whereas these studies have in fact overcome some typical methodological problems of secondary analyses—namely, flaws in standardization (comparability), replicability or reliability (too few items)—this comes at the price of serious theoretical drawbacks.

Gottfredson and Hirschi identify an individual difference variable as the main cause of crime, which is seen as ‘an enduring tendency [that] is well within the meaning of a “personality trait”’ (1990: 109), and they offer a descriptive portrait of the offender in terms of personality traits. It is therefore tempting to equate the derivation of the causal variable with the description of the offender. The authors themselves may have encouraged such an interpretation by labeling their explanatory trait ‘self-control’, and listing the descriptive details under the heading ‘elements of self-control’ (1990: 89–91). I regard this as the fundamental misunderstanding in current operationalizations of the theory. Almost all instruments designed to measure self-control are based on this misinterpretation, at least to some degree. However, degrees can make a difference. What I called the ‘fundamental misunderstanding’ has several distinguishable aspects, and not all instruments rely on all of them. I would counter the various facets of the fundamental misunderstanding by three fundamental conclusions drawn from my interpretation of the theory.

**Conclusion (a): self-control has no elements at all**

Gottfredson and Hirschi base their theory on an analysis of the very behavior they wanted to explain, crime. They identified a number of
features that most criminal, and many other (‘analogous’) acts have in common. The authors underscore throughout their recent writings that there is but one defining characteristic of the acts they discuss: ‘They carry for the individual involved in them the risk of long term negative consequences’ (Hirschi and Gottfredson, 1995: 139). Following this descriptive part, they derive the main explanation from two of three ‘general facts’ (Gottfredson and Hirschi, 1990: 253), drawn from decades of empirical research on criminality: the tendency of all these behaviors to be positively correlated among each other (versatility), and to be positively correlated over time (stability) (see Hirschi and Gottfredson, 1994: 2–3, for a brief and clear summary of this point). Since the constant element across all points of correlated measurements is the individual, and most likely nothing else than the individual, the reason for the observed covariances should be found within the individual personality. This means that there exists some kind of trait. The nature of this trait is then defined by directly translating the common feature of behavior into a personality concept. Perhaps the clearest and most straightforward wording is given in Hirschi and Gottfredson (1994: 3; similar passages may be found at several places in the original formulation of the theory), where they define self-control as: ‘the tendency to avoid acts whose long term costs exceed their momentary advantages’.

It has been subject to some debate whether the trait of self-control can be measured independently of the behaviors it is meant to explain, culminating in the charge of tautology (Akers, 1991, among others). At this point, it is important to draw the conceptual distinction between a trait and behavior. A trait is an abstract characteristic of the person possessing it—and of nothing else—whereas behavior as its observable manifestation is always the result of some complex of determinants located in different sources: traits causing all behaviors within the behavioral class in question (in this case, self-control), other traits that may cause only specific behaviors within but perhaps also outside the class of behaviors, and circumstances of the single situations involved as well as possible interactions between all variables (for a prominent criminological application of these well-known statistical principles prior to the GTC, see Osgood et al., 1988). This means that the behavior as a whole could hardly be expected to be proven unidimensional in empirical investigations. However, if we are interested in measuring self-control as the common cause of all behaviors in question, this general trait should indeed be unidimensional. Put differently, we want to isolate the common trait variance (or ‘true variance’, in terms of psychological test theory) from the variance unique to only some acts within the domain (specific variance) as well as from the random variance specific to the single situation (error variance). That is, we are in pursuit of the general factor in the class of behaviors because this is self-control, and all else is nuisance or measurement error.

However, we are faced with the fact that there is a paragraph entitled ‘elements of self-control’ in the original formulation of the theory which, unfortunately, reads like a cookbook for the development of a self-control
measure. Because by far the most popular instrument in empirical tests of the GTC to date (Grasmick et al., 1993; for published applications see the Appendix in Pratt and Cullen, 2000, as well as many of the more recent studies cited in the introduction above) is thoroughly modeled after this passage, it is essential to understand its logical inconsistency with the basic idea of the GTC. In the respective section, Gottfredson and Hirschi list a number of—again—features of criminal acts (e.g. immediacy of gratification, ease of access to the desired ends, thrill of the acts’ execution and so forth) all or some of them possess, and from that description of the acts, they derive specific traits that match every single behavioral characteristic. This way of reasoning is different from the way self-control was originally derived. Where self-control is nothing else than a general latent factor behind the empirical relationships between all acts in question, the ‘elements of self-control’ are translations of single behavioral characteristics, not all of which must be present in all acts at once, into single personality traits that are logically linked only to the specific feature of behavior at hand. This clearly refers to the specific level below the general one the GTC is usually concerned with.

This conceptual incongruence could immediately lead into serious theoretical inconsistency when rendered into a measurement concept. Gottfredson and Hirschi state that ‘there is considerable tendency for these traits [impulsiveness, insensitivity, physicality, risk or adventure seeking, short-sightedness, non-verbality] to come together in the same people’, and they ‘consider them as comprising a stable construct’ (1990: 90–1). This contention is at odds with decades of research on the structure of personality. In the currently most agreed upon taxonomy of personality traits, the ‘five-factor model’ (FFM) or ‘big five’ model, personality is described in terms of five broad factors which capture the common variance of most more narrowly defined descriptions of human character (for reviews see, for example, Digman, 1990; John, 1990; Wiggins, 1996). On the other hand, the five broad factors are believed to be largely independent of each other. In terms of this model, insensitivity would fall under the taxon of agreeableness, preference for non-verbal and non-intellectual activities under the trait of intellectance (or openness to experience, not to be mixed up with intelligence), a lack of persistence would be an indicator of low conscientiousness, adventure or sensation seeking is best seen as a facet of extraversion and impulsiveness may be loaded by neuroticism as well as low conscientiousness. In other words, included in the list are traits that may belong to every single one of the five FFM dimensions. Even if these narrow traits are not perfectly orthogonal, it seems extremely unlikely that a trait defined as unidimensional happens to become decomposed into elements that are best described by all five factors derived from orthogonal rotation of the entire list of trait adjectives (this is what Marcus (2003) found for the Grasmick et al. scale). It follows that either the FFM lacks validity or the Grasmick et al. scale lacks unidimensionality. There are countless studies demonstrating the generalizability of the FFM across
different cultures, rating methods and sources, statistical procedures, etc. (see the reviews cited above), whereas results on the dimensionality of Grasmick et al.’s and similar scales are at best inconclusive (Longshore et al., 1996; Cochran et al., 1998; Piquero and Rosay, 1998; Arneklev et al., 1999; Piquero et al., 2000; Vazsonyi et al., 2001; Marcus, 2003). Therefore, it seems reasonable to conclude that the latter explanation has empirical evidence on its side. This alone is a serious challenge of the measures’ construct validity, because self-control is conceptualized as one general factor, not as a composite of several more or less independent primary factors. However, as will be shown next, these elements by themselves are in large parts logically inconsistent with the concept of self-control.

Conclusion (b): self-control is a trait of differential control, not motivation

Most if not all of the traits discussed above have been shown to correlate with criminal acts. At least some of them are well founded in psychological theories of personality. Thus, we have a relatively precise understanding of the logical links between these traits and deviant behavior. These links are in some cases completely incompatible with the logic of the GTC, as may best be illustrated with one example out of the list of elements, alternatively labeled risk, thrill, adventure, excitement or sensation seeking. There is considerable empirical evidence that people prone to seek risky or stimulating activities are also prone to commit criminal and analogous acts (Lösel, 1975; Pérez and Torrubia, 1985; Eysenck and Eysenck, 1992; Wills et al., 1995, 1998; Wood et al., 1995; Gordon and Caltabiano, 1996; Lalumière and Quinsey, 1996; Turrisi et al., 1997; not citing the studies with the Grasmick et al. questionnaire). This is no surprise since criminal activities involve some risk and are usually exciting, as Gottfredson and Hirschi do not forget to point out. By close inspection, however, it is easily shown that this trait–behavior connection competes rather with than resides in the idea of self-control.

One basic assumption of Gottfredson and Hirschi is that the desired outcomes of deviant and analogous acts (e.g. money, goods, sex, a state of high, relief from some momentary irritation) are attractive to anyone. That is, in the GTC, offenders do not commit crimes because they value their benefits more than other people do but because they are not sufficiently restrained by the negative consequences these behaviors have for themselves in the long run. The explanation has to be sought in differential control over universal impulses. However, in the case of ‘thrill’ as an outcome of behavior, there are substantial differences between persons in the way they evaluate it. We may all be interested in making money without effort but we certainly do not share the same attitude toward bungee jumping as a leisure-time activity. Some of us may find thrills like this anything but attractive. It is precisely this kind of individual difference that is described by the trait of sensation seeking. Sensation seeking, in its most
elaborated theoretical formulation (Zuckerman, 1994), is a trait of differential motivation to seek or avoid certain situations, caused by differential levels of cortical arousal. The idea of differential motivation as an explanation for crime is one that the authors of the GTC explicitly reject (see also Hirschi and Gottfredson, 1990, 1993). Therefore, there is a theoretical gap between sensation seeking and self-control that cannot be bridged by empirical means. The inclusion of one trait as part of the other (as in the Grasmick et al. measure), or even the equation of self-control with preferences for risk (e.g. Keane et al., 1993; Lamnek, 1994) makes one or both concepts meaningless.

Nevertheless, we should expect sensation seeking and low self-control to be positively correlated in empirical investigations. This is simply the consequence of the fact that many activities, and most prominently criminal ones, are both thrilling and promise immediate gratification at the expense of long-term negative consequences.

Examples for conceptual incongruence may also be found for other traits included in the list of ‘elements’: a preference for physical or easy tasks over intellectual or complex ones implies that the former activities are perceived as more attractive relative to the latter ones—again, a difference in motivation, not control. One who is insensitive toward other persons’ feelings may commit crime because he or she places low value on the immediate negative consequences crime has for others. This insensitive person may nevertheless avoid hurting somebody because he or she fears revenge, that is, has enough self-control to take the long-term negative consequences for him- or herself into account. Even though the elements of temper and impulsivity are more consistent with the idea of control, these examples show that the elementaristic approach to the measurement of self-control is seriously misleading as long as the conceptual foundations of self-control and its apparent elements are as divergent as is the case for at least four out of six traits listed. Concerns regarding their lack of unidimensionality, or their differential validity (Arneklev et al., 1993) cause no trouble for the GTC because they simply fail to have any conceptual underpinnings in this theory and can therefore neither support nor falsify it. Moreover, it is noteworthy that the discussion above is purely logical, and that the conceptual differences between self-control and the traits referred to as elements cannot be overcome by an empirical demonstration that they predict crime or analogous acts. They may do so, but for reasons different from the idea of self-control. I will now turn to some more methodological aspects.

**Conclusion (c): self-control is manifested in behavior, not self-reflection or evaluation**

Not all existing measures of self-control place such an emphasis on the complete and balanced inclusion of all the elements listed by Gottfredson and Hirschi. For instance, Burton and associates (Burton et al., 1994, 1998;
Evans et al., 1997) developed a 12-item scale that has no subscales to enclose all facets but the authors still based their construction on the list of elements (resulting in a low internal consistency reliability). Similarly, Gibbs and Giever (1995; Gibbs et al., 1998) provide a 40-item questionnaire that includes a number of items tapping into most of the elements. The latter instrument, however, is different from the others with respect to the way the authors interpret self-control.

Although Gibbs and Giever covered several aspects as well, they explicitly state that they intended to construct a ‘unitary measure of self-control’ (1995: 244) and seem to follow this statement more consistently than others (e.g. Burton and colleagues; Cochran et al., 1998). The data they present support the homogeneity of their instrument to a much greater extent than, for instance, that of Grasmick et al. Their measure emphasizes one element out of the list which may best be seen as ‘here-and-now orientation’ (see Table III in Gibbs and Giever, 1995, indicating that items like ‘Don’t postpone until tomorrow a good thing that can be had today’, or ‘Take your pleasure where and when you can get it’ are marker variables for the scale). No doubt, this parallels much closer the definition of self-control adopted throughout this article than constructs like sensation seeking or physicality.

So, is this the measure that holds promise for operationalizing self-control in the way it was defined in the GTC? I think not. As I tried to show earlier, this latent trait is directly derived from common variance between acts that share particular features but do not share others. I went on by saying that self-control is nothing else than this covariance between the acts that comprise it. All else is error variance and should be eliminated as far as possible. From this assumption, it follows that attitudinal items may serve as indicators for self-control to the degree they preserve the true individual differences in the behavior of interest, and not add irrelevant or distortional components that detract from true variance.

All instruments discussed so far, including that by Gibbs and Giever, predominantly rely on global descriptions or evaluations of the self or external objects as single indicators (items). At the item level, such abstractions may have superior psychometrical properties compared to single behavioral incidents, since they are in themselves implicit aggregations and, thus, more reliable. Gibbs and Giever rightly point out that the occurrence of observable behavior may not be attributed to the underlying variable of interest alone—varying circumstances may cause substantial behavioral variation from one situation to another. However, what applies to a single act may not apply to aggregations of such acts. It is a fundamental statistical principle that unsystematic errors tend to cancel each other out in aggregates (see Rushton et al., 1983, for prominent applications). Therefore, the advantage in reliability for global, abstract self-assessments decreases as scale length increases.

On the other hand, attitudinal items have some important drawbacks for the measurement of self-control compared to behavioral acts. First, the way...
from actual behavior to questionnaire responses is a long and indirect one. A person is forced to think of an abstract target of attitude or of his or her behavior in the past, evaluate or summarize it, apply some kind of social comparison to find out whether the statement characterizes him or her better than others, and so forth. The more cognitive operations like these are required, the less likely it becomes that the response gives an undistorted picture of behavior (see Henry et al., 1994). Second, these possible distortions can be expected to affect outcomes systematically. Items like those described require some kind of introspection and mostly a reliable—i.e. realistic—assessment of the self, including the consequences of one’s behavior in the past. These are precisely the skills or qualities we would not ascribe to a person low in self-control. Translated into more technical terms, we would expect an ‘attitudinal’ or self-reflective scale for self-control to yield more reliable—and thus valid—scores at the high pole than at the low pole of the trait it measures. This is an undesirable property, especially when we are mainly interested in those people lacking self-control, and there are no means to control for it post-hoc. The obvious alternative is to select the items directly out of the universe of self-control behaviors.

Given that these conclusions about the measurement of self-control hold, we are apparently left with few alternatives to measuring the independent variable just like the dependent one in construct-valid tests of the theory. We are then again faced with the charge of tautology. This may look like a dilemma. However, as I hope to have shown in the preceding section, self-control (the independent variable) and crime (one possible dependent variable) are not at all ‘one and the same thing’ (Akers, 1991: 204) from a theoretical point of view. Low self-control is seen as one general cause of crime, but not the only one. Crime can be seen as one manifestation of low self-control, but not the only one. Akers’ own discussion of the tautology argument continues with the conclusion that ‘to avoid the tautology problem, independent indicators of self-control are needed’ (1991: 204). That is, the problem is one of measurement, not theory. Fortunately, indicators of self-control independent of crime are easily identified. Gottfredson and Hirschi enumerate a respectable list of analogous acts in various chapters of their book as well as in later writings (e.g. Hirschi and Gottfredson, 1994: 9; see also there for a reply to Akers). If these acts, measured appropriately, are not crime (which is obvious) but predictive of crime, if they comprise a general latent factor (and not many), and if they produce reliable variance across diverse populations (within criminal and non-criminal samples, respectively, and between them) and cultures not affecting their relationships to crime or other dependent variables of interest, this would support the theory. Otherwise, the theory or important parts of it must be regarded as falsified. A theory which is falsifiable cannot at the same time be tautological.

There are some approaches which go half way into that direction. For instance, in several studies (e.g. Arneklev et al., 1993; Wood et al., 1993;
Piquero and Tibbetts, 1996; Burton et al., 1998), measures of ‘imprudent behavior’, minor deviance and the like are included. Ironically, the respective scales were interpreted as dependent or control variables, whereas they may be judged much better indicators of the independent variable than the self-control scales used therein.

One exception to this rule is the article by Evans et al. (1997) who considered their imprudent behaviors scale as both a dependent and as a possible independent variable. Although this may again be judged a substantial progress, I still see considerable room for improvements. Put briefly, I regard this scale as too short to be reliable as a trait measure (it has 18 items), too heavily loaded by illegal activities to be judged independent of crime (it includes five illegal drug use items along with several other activities prohibited by the law), not broad enough to test a general factor hypothesis and possessing inadequate scale anchors for the time frame measured (see section below). Similarly, but even less inclusive, Gibbs and Giever (1995), and LaGrange and Silverman (1999) also developed additional, behavioral measures of self-control, covering two kinds of acts with five and two items, respectively. All but one objection I raised against Evans et al.’s scale apply to these measures to an even greater extent (the exception is the charge of not being independent of crime).

Finally, Herbert (1997) introduced a 24-item behavioral instrument in a doctoral dissertation. However, her scale includes 11 items tapping into volitional acts (academic misconduct, legal drug use and misconduct toward parents) along with 13 items asking for involvement in accidents, or victimization by others—incidents which happen to a person rather than being actively conducted. By definition, these events are not under complete volitional control of the person they happen to (only few people intentionally involve themselves in accidents). They may be caused by low self-control (one plausible explanation for the well-known positive correlation between these events and offending), but they may also be caused simply by bad luck (that is, random variance). In terms of measurement theory, this adds more error variance to the scale than necessary. Thus, accidents and victimization may be seen as second-best indicators for self-control at best, but more adequately as dependent variables.

Requirements for a construct-valid self-control measure

So far, this article largely dealt with how self-control should not be measured. Of course, identifying a problem is usually easier than solving it. But from the discussion above, there also follow a number of implications on how an appropriate instrument should look. In this section, I refer to these aspects as requirements for a self-control measure.

Requirement (a) follows directly from the inherent problems of secondary analyses and claims that findings from tests of a theory should be
replicable and comparable to one another. Just like this demand, an easy and cost-effective solution to it is not entirely novel: a completely standardized self-report questionnaire like those proposed by Grasmick et al. and all researchers who developed self-control scales in more recent years.

The second requirement \((b)\) simply states that individual items should ask for overt behavioral incidents, not introspective self-assessments, attitudes or external events. This conclusion is based on the discussion above where I tried to develop the general idea that all of the latter types of item are likely to add irrelevant variance to the measure and, thus, are at least inferior if not improper indicators for self-control.

Requirement \((c)\) is an extension of requirement \((b)\) in that it makes explicit which kinds of behavior may be considered as adequate for inclusion. This should set the margins for a possible ‘universe’ of indicators. As follows from the discussion on how self-control is defined—and how it is not—appropriate acts should primarily offer some kind of short-term gratification, attractive to anybody, at the expense of possible long-term negative consequences outweighing the benefit. A carefully constructed measure should also try to take some additional features into account (obvious benefits and execution; no particular training, little effort, skill and planning required) which Hirschi and Gottfredson regard as also common but ‘less defining’ (1995: 139) for acts of low self-control. Examination of this requirement would most likely lead to the rejection of self-reflective items, because these may match the defining characteristic of self-control only if (a) they tap into a trait consistent with the idea of control (true for temper and impulsivity only), and (b) responses are realistic cognitive translations from behavior into self-concepts related to such traits.

The next requirement \((d)\) is as straightforwardly derived as the preceding ones but a little more difficult to translate into action. From the fact that the ‘traitness’ of self-control is based on the empirical findings of versatility and stability, it follows that a comprehensive measure should take these two aspects into account. That is, first, the generality of the trait can only be retained in an instrument if we assure that it captures the essence of the class of acts as a whole and not merely of one or two specific subcategories like, say, alcohol consumption. The logic of self-reflective measurement as an implicit aggregation of prior conduct has an inherent means to handle this specific problem, but none of the existing behavioral measures had approached it systematically. This would require a sufficiently large number of sufficiently diverse indicators—diverse with respect to anything but the defining features of self-control. Such an operationalization would allow for an empirical test of the versatility (general factor) hypothesis characterizing the concept of self-control. This is also one reason why an adequate measure of self-control must contain a relatively large number of items, though it does not have to be comprehensive with respect to the entire behavioral domain.

With regard to the second part of requirement \((d)\), stability, fulfillment
may be empirically tested in longitudinal designs (Arneklev et al., 1999). An easier and more cost-effective alternative is to apply a retrospective measurement approach. Important for the concept of self-control is that it should cover a long period in life (this would preclude the use of longitudinal designs for most practical purposes) that is constant across all respondents. With a few exceptions in secondary analyses (e.g. Polakowski, 1994), previous measures had either used a time frame of one year or less or ignored the issue altogether (the Grasmick et al. items, for example, are all worded with respect to present time).

Closely connected with the latter point in particular, and with requirement (d) in general, is the next requirement (e). It states that systematic influences on behavior independent of self-control should be eliminated to the greatest possible extent (taking cost-effectiveness into account). The GTC knows only two such variables: opportunity and age. Fortunately, both can be controlled by the same means as the versatility and stability issues.

Opportunity is largely seen as a random factor specific to the concrete situation (for an exception, see the discussion on white-collar crime in Gottfredson and Hirschi, 1990 and Hirschi and Gottfredson, 1993). Having a wide range of different types of acts in the measure, and covering decades of the life-time with naturally varying circumstances should eliminate these influences to a large extent. As Hirschi and Gottfredson put it, ‘self-control and opportunity may [...] interact for specific crimes, but are in the general case independent’ (1993: 50). This is merely an application of the familiar statistical principle that measurement errors tend to cancel each other out in composites. The more different items are aggregated, the better this principle works—yet another argument for a longer scale.

Age is also an ‘important cause of crime that self-control cannot explain’ (Hirschi and Gottfredson, 1993: 50). It follows that age effects on crime should always be controlled for in tests of the GTC. This may be done by means of statistical partialing in every individual study (e.g. Evans et al., 1997, among several others). This is in fact a necessity when one uses a scale format that has respondents indicate their behavior during a particular period before the time of measurement, for instance during the last year (all behavioral scales discussed above have anchors like this). In this case, age differences in crime—which Gottfredson and Hirschi explicitly explain by natural maturation, not self-control—could seriously distort results if not controlled for. However, one consequence of this format for behavioral self-control scales that cannot be removed by partialing—which has an effect only on correlational statistics—is the fact that test scores cannot be directly compared between subjects of different age across studies unless age-specific norms are established—a laborious and costly task. For this reason, a different strategy is recommended here. A retrospective format enables all participants to reply to questions about the same period in their lives. This is an application of a well-known alternative to partialing for controlling a variable: holding that variable
constant. The worst thing that can happen is that this procedure does not work (that is, the measure is correlated with age). In this case, however, partialing is still an available option so that nothing is lost.

The next requirement (f) is a consideration of the tautology argument which eventually turned out to be merely the demand for a self-control measure logically independent of crime. Again, it would be fortunate to have a respectable number of different types of behavior in the scale at least some of which cannot be judged as criminal or even deviant (e.g. wasting money). If legal and illegal acts are measured in the same study, the issue of equivalence between criminal and analogous acts may also be tested empirically.

The basis for my final requirement (g) has only been touched upon so far. The most general idea behind the general theory is its generality, or the invariance of the explanatory mechanisms over varying circumstances, kinds of offenses, populations, cultures, historical periods and so forth. One essential part of this ‘general generality’ is that lack of self-control is not seen as a rare defect that explains the high-frequency offending for a small group of sociopaths (I explicitly disagree with Cohen and Vila, 1996, on this point). On the contrary, self-control is conceptualized as a continuously distributed personality trait possessed by everybody to a differential extent. Therefore, it would be desirable to have a measure that could be used with every kind of population in every culture of the world. A strict realization of this requirement, however, may be beyond the limited potential of psychometric instruments, but the goal of general applicability may at least be approached to some extent. This would require a mixture of acts with varying degrees of seriousness with respect to the negative consequences, in order to produce reliable variance at different levels of self-control. In addition, care should be taken that the majority of people living in different cultures would have access to situations where the acts of low self-control may occur. That is, everyday behaviors like arguments with partners, wastefulness or school problems (provided compulsory education) are preferable over acts tied to more specific contexts.

Not all of the above requirements are equally important. I regard requirements (b) through (d) as the crucial points for a theoretically adequate operationalization. With (a), (e) and, in a certain sense, (f), I added some more technical but still highly important requirements for a psychometrically sound instrument. The latter point (g) is more a desired than an absolutely necessary category, fulfillment of which may be a matter of degree.

Conclusion

Table 1 summarizes the seven requirements and the way they may be approached in constructing a self-control scale. Also included in the Table are judgments of the extent to which each requirement has been met by the
three main classes of measures employed in previous investigations. It seems obvious that all of these approaches at least partially fulfill some requirements but also clearly fall short of satisfying other important needs. It is hard to draw general conclusions on the highly diverse class of secondary analyses, apart from their lack of replicability. Attitudinal scales typically yield a psychometrically sound measurement, but this strength cannot overcome the fundamental problem of theoretical inadequacy. If an instrument does not match the definition of the construct it purports to measure, it will logically fail to deliver valid tests of any relationships of that construct (though it may well be very useful for different purposes). Behavioral imprudence scales, on the other hand, come much closer to the theoretical concept of self-control but appear to fall short of meeting acceptable standards of psychometrics. In all cases, self-control measurement is likely to be seriously impaired for one or the other reason. This has serious consequences for interpreting the current empirical status of the GTC.

In their meta-analysis, Pratt and Cullen (2000) estimated the mean correlation between self-control and dependent variables to be $r = .26$ for attitudinal, and $r = .28$ for behavioral (reanalyses plus imprudence scales) measures of self-control, respectively. These figures do not appear to differ substantially and both appear to point to a moderate, though significant relationship between self-control and criminal and analogous behaviors. In addition, rival theories were also tested in the same meta-analysis and found to produce slightly lower but still significant effect sizes. Thus, the world seems to be safe for both proponents and opponents of the GTC, leaving enough room for alternative operationalizations of self-control, according to one’s personal preference, for alternative explanations of crime, and for an ongoing debate on the merits and shortcomings of the GTC. This picture of moderate support for any position is at odds with Gottfredson and Hirschi’s original claim of presenting a general theory of crime but does not justify dismissing the theory altogether.

If there are serious flaws in the measurement of self-control, however, the picture may look completely different. When self-control has never been adequately measured, we simply do not know the status of the GTC. We will therefore not be able to draw firm conclusions on any of the controversies surrounding the theory. Some evidence on how far previous results may differ from more psychometrically sound ones comes from research employing an alternative self-control scale, designed to meet all seven measurement requirements (Marcus, 2003). This instrument, labeled Retrospective Behavioral Self-control Scale (RBS), was found to measure something substantially different than the Grasmick et al. scale, both in terms of internal structure and relationships to other personality constructs. If correlated with deviant workplace behaviors as dependent variable, the RBS produced an effect size of $r = .63$, far beyond Pratt and Cullen’s upper bounds of the 95 percent confidence intervals for behavioral (.40) or attitudinal (.32) measures (Marcus et al., 2002). Self-control alone
Table 1. Requirements for a self-control scale and their realization in different kinds of measure

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Recommended means for realization</th>
<th>Extent of realization in different measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SA</td>
</tr>
<tr>
<td>(a) Replicability, comparability</td>
<td>Standardized test</td>
<td>no</td>
</tr>
<tr>
<td>(b) Behavioral basis</td>
<td>Assessment of volitional behavioral acts</td>
<td>partially</td>
</tr>
<tr>
<td>(c) Match with the definition of self-control</td>
<td>Assessment of acts that provide immediate gratification at the expense of long-term negative consequences outweighing the benefit</td>
<td>partially</td>
</tr>
<tr>
<td>(d) Consideration of versatility and stability</td>
<td>Versatility: Inclusion of several categories with highly diverse content Stability: Retrospective measurement covering long periods in life</td>
<td>rarely</td>
</tr>
<tr>
<td>(e) Elimination of systematic influences beyond self-control (opportunity, age)</td>
<td>Opportunity: Diversity of content and length of time frame covered should cancel out systematic influences Age: Holding age constant via scale format</td>
<td>rarely</td>
</tr>
<tr>
<td>(f) Independence of crime (non-tautological measurement)</td>
<td>Items tapping into non-criminal, ‘analogous’ acts (some minor offenses may be included for comparisons)</td>
<td>mostly</td>
</tr>
<tr>
<td>(g) Broad coverage across different populations and cultures</td>
<td>Behavior that most people have access to; widespread acts</td>
<td>not addressed but possible</td>
</tr>
</tbody>
</table>

Note: SA: secondary analyses; AS: attitudinal scales; IBS: imprudent behavior scales.
accounted for about the same proportion of variance in the dependent variable as an entire set of 23 alternative independent variables (Multiple R = .65) (Marcus and Schuler, forthcoming). Thus, our judgment of self-control theory may change dramatically once self-control is properly measured.

An alternative interpretation of Pratt and Cullen’s findings, for example, may be that the robust relationship between delinquency and attitudinal measures of the ‘elements of self-control’ may point to the fact that there are a number of robust predictors of crime within the personality sphere beyond self-control. Whether these findings hold once self-control is accounted for would still be a matter of additional research as are Pratt and Cullen’s results on the merits of rival theories. When the results reported by Marcus et al. are to be found to generalize to other settings, such incremental validities may eventually turn out to be artifacts of imperfect measurement of self-control. Similarly, the meta-analytic finding of almost identical validities for both attitudinal and (single-indicator) behavioral measures of self-control may not point to the insensitivity of self-control to different operationalizations, as Pratt and Cullen concluded, but rather to the coincidence of finding the same incorrect effect sizes with two different types of flawed measurement. With the present state of affairs, we simply do not have sufficient information to give definite answers to these questions.

The studies summarized by Pratt and Cullen addressed mainly the issue of criterion-related validity of the GTC. With an adequate measure of self-control, however, we may also be able to empirically bolster the sometimes speculative arguments in discussions surrounding more conceptual issues. For instance, if self-control is more than old wine in new bottles, a measure of this construct should display discriminant validity if related to measures of more traditional personality traits, and incremental validity if related to independent variables. Without such a measure, the conceptual standing of the GTC within the multitude of theoretical traditions in criminology is just as hardly tested as is its value for predicting crime.

So in conclusion, there may be little gain in understanding and evaluating the GTC unless the measurement problem has been solved. Given the deficiencies in prior research with respect to this task, it seems questionable whether there is anything we know for sure about the theory after more than 10 years of intensive research—what a waste of time!

Notes

This research was supported by grants from the Deutsche Forschungsgemeinschaft (SCHU 422/9–2) and from the Bundesland Baden-Württemberg. I am also grateful to Stefan Höft and Dick Weissert for comments on earlier drafts of this article and to Travis Hirschi for his enduring encouragement.
1. Some elements of the theory (e.g. the invariance hypotheses of sex or age distributions of crime) may be tested without measuring self-control (e.g. Jang and Krohn, 1995). These types of investigation do not belong to the topic of this article. On the other hand, self-control may also be conceptualized as a dependent variable if its causes (e.g. parental management) are at the core of interest. The measurement issues discussed here apply equally well to such studies.

2. This interpretation contradicts the idea that the variance common to criminal and analogous acts can be investigated after self-control has been removed from that relationship (Paternoster and Brame, 1998). Self-control basically is this covariance.

3. For a general discussion of the controversies around the FFM, see the Block (1995) vs Costa and McCrae (1995)/Goldberg and Saucier (1995) debate. For our purpose, suffice it to say that the FFM has the most robust empirical basis of all factor-analytic approaches to personality to date but its status as the ‘final word’ on personality theory is still a matter of controversy.

4. A comprehensive technical treatment of the varying findings in confirmatory dimensionality tests of Grasmick et al.’s scale is beyond the scope of this article. Briefly, the different authors employed different sample sizes, different fit indices and—most importantly—different models to test for unidimensionality. All these differences are likely to contribute to variations in model fit. The three most recent studies cited provide an adequate baseline model (a model with correlated first-order factors) against which to compare the fit of a one-factor solution, and all found a substantial or even dramatic drop in fit for their models with one general factor.


6. I do not share Evans et al.’s opinion that Gottfredson and Hirschi distinguish between criminal and non-criminal yet illegal ‘analogous’ acts conceptually. In my view, there is no conceptual distinction between criminal and analogous acts in the GTC at all. The best and easiest way to find independent indicators is to choose acts which have long-term negative consequences other than criminal conviction.

7. Hirschi and Gottfredson (1993; see also Longshore et al., 1996) seem to contrast self-reports and behavioral measures as if they were opposites. This, I think, is not true. A measure can be a self-report and still contain only behavior. More meaningful distinctions may be drawn between self-reports and observational methods, on one hand, and between concrete acts and abstract reflections, on the other. I would contend that it is the latter Hirschi and Gottfredson dismiss, not self-reports per se.

8. It is noteworthy that Evans et al. also controlled for gender in their study. This, I think, is a conceptual mistake because, in the GTC, gender differences in crime and analogous acts are indeed explained by substantial differences in self-control between sexes. Thus, Evans et al. may have partialed true trait variance out of their relationships.
9. Although there are numerous problems with retrospective measurement if one is interested in exact estimates of the frequency of events, the method has been found to be quite accurate and more so than in areas of self-reflection when it is used to rank-order individuals on measures of delinquency (Henry et al., 1994).

References


Eysenck, Hans-怎么能确认段落的完整性？段落为英文，包含多个作者和他们的作品。我不确定这个段落是否完整，因为它似乎是一个引文列表或参考文献。它没有提供具体的上下文或文章内容。你需要提供更多的上下文信息才能更好地理解这个段落。


BERND MARCUS received his doctoral degree in business administration and economics from the University of Hohenheim and is currently Research Assistant (Wissenschaftlicher Mitarbeiter) at the Chemnitz University of Technology, Germany. Recent publications are on deviant behavior in the workplace, personality theory and measurement and personnel selection and performance.