

The Theory of Planned Behavior Without Compatibility? Beyond Method Bias and Past Trivial Associations¹

Florian G. Kaiser², *Eindhoven University of Technology Eindhoven, The Netherlands*
P. Wesley Schultz, *California State University, San Marcos*
Hannah Scheuthle, *Swiss Federal Institute of Technology Zürich, Switzerland*

Overreliance on one measurement approach can challenge accurate statements about reality, as findings can represent by-products of the compulsory measurement paradigm. Within the theory of planned behavior (TPB), the compatibility principle represents one such strictly imposed paradigm. Using 2 cross-sectional surveys of 1,394 volunteers and involving structural equation models, we demonstrate that the widely employed practice of measuring TPB constructs is confounded with method-implied bias. This means the theory cannot conclusively reveal origins of a behavior. Our results also suggest that on an aggregated level, when method bias is eliminated, its constructs are linked in hypothesized ways. Adopting a more general model—thus, adopting a more traitlike conceptualization of attitudes—has interesting implications for social psychology and its current trends.

This is the peer reviewed version of the following article: Kaiser, F., Schultz, P. W., & Scheuthle, H. (2007). The theory of planned behavior without compatibility? Beyond method bias and past trivial associations. *Journal of Applied Social Psychology*, 37, 1522-1544. <http://dx.doi.org/10.1111/j.1559-1816.2007.00225.x>, which has been published in final form at <http://dx.doi.org/10.1111/j.1559-1816.2007.00225.x> This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

¹ This research was supported by Grant TH-15/99-1 from the Swiss Federal Institute of Technology at Zürich, Zürich, Switzerland; and by the Department of Technology Management at Eindhoven University of Technology, Eindhoven, The Netherlands. We gratefully acknowledge Rainer Guski (Ruhr University, Bochum, Germany) and Volker Linneweber (Otto-von-Guericke University, Magdeburg, Germany) for their institutional support of the data collection. We thank Gundula Hübner, Marcel Hunecke, Bernd Six, and two anonymous reviewers for their comments on earlier drafts of this paper; Steven Ralston for his language support; and the volunteers who completed questionnaires.

² Correspondence concerning this article should be addressed to Florian G. Kaiser, Eindhoven University of Technology (IPO 1.22), P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands. E-mail: f.g.kaiser@tue.nl

Sexual assaults, hate crimes, overconsumption, traffic congestion, and urban poverty are some of the many problems faced by modern societies. Psychology, as the science of human behavior, can help us to understand and solve these problems (Oskamp & Schultz, 1998). Indeed, changing behavior in response to a problem is the ultimate objective of applied psychology. However, empirical research can only contribute to the solution of societal problems if its predictions about the origins of behavior are accurate conclusions about reality, and more than just by-products of the employed measurement paradigm.

One of the most widely endorsed behavior explanation models in applied social psychology is the theory of planned behavior (cf. Ajzen, 2001). The theory has been employed across a number of behavioral domains, including condom use, alcohol consumption, seatbelt use, water conservation, public transportation use, and recycling, to name but a few (Armitage & Conner, 2001). Because of the way this theory addresses influences external to a person (e.g., opportunity, time pressure), it is applied almost exclusively in a specific manner, to explain a single action (e.g., car oil recycling, being a member of an environmental organization; Kaiser & Gutscher, 2003). However, such specific, often single-item measures are either relatively unreliable or frequently not valid in comprehensively capturing the essence of more general attitudinal dispositions (e.g., Ajzen & Fishbein, 2005).

Compliance with a strict measurement paradigm, the so-called *compatibility principle* (originally called *correspondence principle*; e.g., Ajzen & Fishbein, 1977), is the traditional way to improve the quality (i.e., reliability) of the concept measures within the theory of planned behavior (Ajzen & Sexton, 1999) and within attitude-behavior research in general (Ajzen & Fishbein, 1977, 2005). Unfortunately, reliance on a particular measurement procedure to assess a concept confounds the concept definition with a compulsory measurement paradigm (cf. Michell, 1986). As a result, research that complies with the compatibility principle then

becomes vulnerable to the criticism that the results merely represent method-implied bias and do not constitute generalizable, valid statements about underlying relationships (e.g., Dawes & Smith, 1985; see also Michell, 1986).

In this paper, we present research in which we deliberately violate the compatibility principle in our tests of the theory of planned behavior. We predict that the theory—and also a close attitude–behavior link—only depends on compatibility at a specific, single-performance level. However, in its general version (see Kaiser & Gutscher, 2003), we expect the theory of planned behavior to retain its explanatory power and validity, even if the compatibility principle is violated.

The Theory of Planned Behavior

According to the theory of planned behavior (TPB), intentions are the best predictor of behavior. Intentions, in turn, are predicted by attitudes, subjective norms, and behavioral control. *Attitudes*, in the TPB, include evaluative beliefs regarding certain outcomes and an estimation of the likelihood that these behavioral outcomes will occur. *Subjective norms* are the strength of normative behavioral beliefs and the motivation to comply with these beliefs. Together with a person's confidence in his or her ability to engage in a certain behavior (i.e., a person's perceived behavioral control) and the perception of the expectations of relevant others (i.e., subjective norms), attitude toward performing a particular act determines a person's intention to perform the behavior in question.

The more a behavior depends on the presence of appropriate circumstances that are external to a person (e.g., opportunities), the less a behavior is intentionally controllable. Thus, in addition to the relationship between perceived control and intention, the TPB also models a direct influence of perceived behavioral control (PBC)—representing actual control—on behavior (Ajzen & Madden, 1986). Evidence has indicated that including this direct PBC–behavior path provides a more complete explanation of behaviors that are beyond a person's

volitional control (e.g., Eagly & Chaiken, 1993; Madden, Ellen, & Ajzen, 1992).

Research using the TPB has found it to be a useful tool for understanding intention and behavior. Usually, the theory is applied to a single behavior clearly defined in terms of action, object, context, and time. For example, “The next time you change the oil in your car, how likely is it that you will return the used oil to its retailer to be recycled?” In an early review of 13 studies, Fishbein and Ajzen (1975) found a mean multiple correlation of .75 for the prediction of intention (see also Ajzen, 1988). More recently, a meta-analysis of 185 TPB studies found a multiple correlation of .63 for prediction of intention (Armitage & Conner, 2001).

With regard to behavior, the theory does a reasonable job as well. Intention tends to be a good predictor of behavior. Reported correlations have ranged from .40 to .95 (e.g., Ajzen, 1988; Ajzen & Fishbein, 1980; Kothandapani, 1971). In their meta-analysis, Armitage and Conner (2001) found a mean correlation of .47 between intention and behavior, and a direct relationship of perceived behavioral control on behavior of .37. When combined, behavioral control and intention explained 27% of the variance in behavior. This is consistent with Ajzen’s (1991) estimate that the TPB can explain 25–30% of the variance in behavior (see also Sheeran & Orbell, 1998). Turning the percentages around, approximately 70–75% of the variance in behavior remains unaccounted for by the theory. One possible source of this large proportion of unexplained variance is unreliable concept measurement, as the TPB must be tested specifically and often is based on single-item measures, which are inherently unreliable.

Specific and Aggregated Realizations

By aggregating across different behaviors from one domain—for example, instead of solely inquiring about car oil recycling, expanding one’s view to paper and glass recycling, car use avoidance, environmental organization membership, and other conservation activities—one eventually arrives at more reliable concept measures (e.g., Epstein, 1983; Kirkpatrick, 1997). Compound measurement, sometimes called the *principle of aggregation* (e.g., Ajzen & Fishbein, 2005), has a longstanding tradition in attitude research as well (e.g., Fishbein & Ajzen, 1974, 1975; Weigel & Newman, 1976).

Unfortunately, aggregation is not an option with the TPB because PBC only works as a proxy for actual control when the circumstances beyond one’s volitional control of a particular behavior are the same for everyone or, at least, can be unequivocally perceived by everyone (Kaiser & Gutscher, 2003). When predicting a unique behavior at a certain time in a particular location, the influences of specific contextual constraints on behavior may well be identified and may be cited as reasons for inaction. For example, a person who lacks recycling bins is unlikely to recycle and will refer to the lack of bins as an excuse. The moment several behaviors (e.g., bike riding, energy conservation, recycling) are combined into a single overall performance measure, though, the person-independent/behavior-specific restrictions and affordances become numerous and heterogeneous. The more diverse and disparate these external influences, the less able we are to collapse them into a single uniform predictor of an individual’s actions. For example, if we attempt to summarize all the affordances and restrictions of a bicycle-friendly environment (e.g., The Netherlands) with a well-established recycling system that does not subsidize home energy conservation, we will probably fail to establish a reasonable estimate of behavioral control. Thus, a compound measure of the self-assessed, control-relevant aspects and influences across a composite of actions, representing a valid summary of the actual control that is left to a person, is unachievable following the principle of

aggregation (for evidence, see Kaiser & Gutscher, 2003).

Fortunately, control-relevant influences external to a person that affect different behaviors and individuals differentially can be directly incorporated into a behavior measure, if this measure is based on the Rasch model (e.g., Kaiser & Wilson, 2004). With this modification in the criterion, the theory otherwise remains unchanged—except for omitting PBC’s direct influence on behavior—when it is applied generally to explain an entire class of behaviors, such as a person’s conservation performance (e.g., Kaiser, Hübner, & Bogner, 2005; Kaiser & Scheuthle, 2003). To our knowledge, this general version of the TPB, originally proposed by Kaiser and Gutscher (2003), is rarely adopted. Instead, the compatibility principle in the measurement of the concepts is the approach usually taken to increase the quality of the attitude and behavior measures, and thus to strengthen the attitude–behavior link.

The Compatibility Principle

According to the compatibility principle, strong correlations between attitude and behavior can only be expected if concepts are compatible in terms of the action involved, the target at which they are directed, and the context and time of their performance (Ajzen & Fishbein, 2005). Technically, the items for the attitudinal and behavioral measures must minimally be on the same level of abstraction; that is, specificity or generality with respect to the addressed behavior, its performance goal, and when the act is expected to occur (Ajzen & Fishbein, 1977, 2005; Eagly & Chaiken, 1993). With this measurement paradigm and particularly with its common implementation, however, the reason for strong correlations between attitudinal and behavioral items could be artificial and is not necessarily substantial.

Survey research has shown that measures with a parallel semantic content or with a high similarity with the criterion often share variance that depends on wording (Lindell & Whitney, 2001). In most instances, such common method variance has the effect of artificially inflating relationships (cf.

Campbell & O'Connell, 1982). For example, remembering an earlier response regarding the importance of conserving something distorts our assessment of the reported engagement frequency. This stems from an order or memory effect, or is an expression of a respondent's wish to appear consistent (cf. Schwarz, 1999).

Item measurement on the same level of specificity, which is stipulated by the compatibility principle in its standard realization, results in a parallel semantic item content, as every item commonly encloses the target behavior. Predictably, preceding answers to questions about, for example, glass recycling frequencies, the importance of glass recycling, the social desirability of glass recycling, and the constraints to glass recycling are likely to bias a person's expressed glass recycling intention (and vice versa), regardless of his or her actual motivation. In other words, repeatedly asking about the same action can artificially exaggerate the strength of associations, as people want to respond in a consistent manner. Note that this common method variance is expected to occur, regardless of the different question-and-response formats that are normally used in the assessment of the different concepts.

Strictly speaking, in its common implementation, the compatibility principle essentially confounds the concept definition and the measurement paradigm (cf. Michell, 1986). Thus, it is theoretically possible that compatible measurement actually exploits common method variance, rather than elucidating true associations between concepts. As such, empirical research based on the compatibility principle cannot provide a clear and unambiguous understanding of the substantive associations among the concepts in question (cf. Dawes & Smith, 1985). Not surprisingly, in its usual implementation, the TPB is sporadically even challenged with psychological irrelevance (cf. Ajzen & Fishbein, 2005).

The study of associations between unobservable, latent concepts, such as attitude, subjective norms, perceived control, and intention, necessarily requires variation in the employed measurement procedure. If variation is impossible, the substance of a conceptual relationship cannot be disentangled from method-induced covariances or other theoretically

irrelevant ones (cf. Eid, 2000). The simplest variation can be achieved when concepts are linked to a class of behaviors or a behavioral domain, such as conservation behavior, rather than to single acts, such as newspaper recycling. Compound measures allow us to aggregate different sets of items for each concept measure, as long as each behavior used in the item formulation is sampled from a single class of activities (cf. Epstein, 1983; Kirkpatrick, 1997).

Research Goals

The questions we ask, at least in part, determine the answers we get (Schwarz, 1999). To explore the possibility of measurement bias in attitude–behavior studies, we must systematically vary our concept measures. The TPB rests on a rigorous measurement paradigm—the compatibility principle—which in its narrow interpretation prohibits such variation. Together with the fact that the theory is typically operationalized specifically toward a single behavior, which additionally weakens the quality of any concept measure, we are at considerable risk of erroneously attributing our findings to substantive, rather than artificial matters. In other words, for the TPB to contribute to an unequivocal understanding of the origins of a behavior, it must be freed from literal compatibility in its concept measures.

In this paper, we test whether the TPB needs to refer to the same behavior in the measurement of its concepts or whether it can be freed from this common implementation of the compatibility principle. Our study aims to provide evidence that the theory retains its explanatory power and validity, even beyond such a parallel construction of its measures, if it is grounded in aggregated measures and, thus, is tested in its general format. This, in turn, requires a new type of behavior measure. Applied specifically and based on relatively unreliable concept measures, the theory depends on the compatibility principle. But when applied generally, based on broad concept measures, the theory can be freed from method-implied common variance without reducing its ability to explain a class of behaviors.

Method

Participants and Procedure

Our samples were taken from resident registers of 11 communities with different numbers of inhabitants in 2000 and 2001. There were seven communities located in the western part of Germany, and four were located in the eastern part. Of the 5,894 randomly selected persons in West Germany (i.e., the first data collection), 607 persons returned completed questionnaires (response rate = 10.3%). Of the 5,000 selected persons in East Germany (i.e., the second data collection), 787 returned completed questionnaires (response rate = 15.7%). Participants' median age was 45.00 years ($M = 45.94$; range = 18 to 89). The proportion of females was 55.5%. While in the East German subsample females were slightly overrepresented, $\chi^2(1, N = 1387) = 4.48, p = .034, \phi^2 = .003$ (58.0% compared to 52.3% in West Germany), no difference was found in the two subsamples' mean age, $F(1, 1381) = 0.56, p = .453, \eta^2 < 0.1\%$.

Given the relatively low response rate, our participants cannot be regarded as representative of the country or the communities in which they live. For the purpose of this study, however, representativity is not necessary. It is sufficient that the participants reflect a wide range of diversity without indication of severe restrictions of range in the attitudinal and behavioral variables that are at stake.

Measures

The questionnaire consists of five measures: attitude toward behavior, subjective norms, PBC, behavior intention, and behavior. All measures conform to common practice in this field of research (e.g., Ajzen & Madden, 1986; Madden et al., 1992). Following previous studies (Kaiser & Scheuthle, 2003; Kaiser et al., 2005), we used two bipolar response scales for each of the four components of the TPB, and each construct was measured for each of the six behaviors.

In the first data collection, the four TPB components were

measured in a compatible way. In this questionnaire, attitude, subjective norms, PBC, and intention were measured with two separate response scales for each construct. There were six specific conservation behaviors: “I bring empty bottles to a recycling bin”; “I collect and recycle used paper”; “I refrain from driving my car in or into the city”; “I have pointed out unecological behavior to someone”; “I buy organically grown produce”; and “I am a member of an environmental organization” (see Appendix).

In the second data collection, we again used two separate response scales to assess each of the four TPB concepts. This time, however, we employed 12 unique conservation behaviors with every TPB component, so that all concepts were measured in an incompatible way. Overall, there were 6 specific conservation behaviors in the compatible measurement, and 48 behaviors (6 specific models X 4 concepts X 2 response scales) in the incompatible measurement (see Appendix). All 48 activities represent conservation behaviors previously recognized as such (see Kaiser & Wilson, 2004).

Attitude. Attitude was measured by rating each of six conservation behaviors on a 5-point scale ranging from 1 (*bad*) to 5 (*good*), and a 5-point scale ranging from 1 (*inappropriate*) to 5 (*appropriate*). The 12 attitude items were internally consistent in both data-collection waves ($\alpha_1 = .77$, $\alpha_2 = .62$).

Subjective norms. Subjective norms were measured by rating each of six behavior statements on a 5-point scale ranging from 1 (*unlikely*) to 5 (*likely*) with the opening phrase “Most people who are important to me think I should . . .,” and on a 5-point scale ranging from 1 (*disagree*) to 5 (*agree*) with the opening phrase “Most people important to me . . .”. The 12 subjective norms items were reasonably internally consistent ($\alpha_1 = .78$, $\alpha_2 = .78$).

Perceived behavioral control. PBC was also measured by rating each of six conservation behaviors on a 5-point scale ranging from 1 (*difficult*) to 5 (*easy*), and on a 5-point scale ranging from 1 (*complicated*) to 5 (*simple*). Internal consistency of the 12 control items was as follows: $\alpha_1 = .62$, $\alpha_2 = .66$.

Behavior intention. Behavior intention was measured by means of six conservation behavior items that were rated on a

5-point scale ranging from 1 (*unlikely*) to 5 (*likely*) with the opening phrase “I will . . . ,” and on a 5-point scale ranging from 1 (*undecided*) to 5 (*decided*) with the opening phrase “I intend to . . .”. Internal consistency of the 12 items was acceptable ($\alpha_1 = .67$, $\alpha_2 = .69$).

Behavior. General behavior was measured in both surveys with 50 conservation behaviors, all of which were found to represent one class of activities in previous research (Kaiser & Wilson, 2004). In other words, the compound of all 50 behaviors can be used as a one-dimensional conservation performance measure (i.e., the General Ecological Behavior [GEB] scale; Kaiser, 1998). Some responses were in a dichotomous, *Yes/No* format (indicating either *ecological engagement* or *nonengagement*; $n = 18$). Others were recoded from a polytomous response format to a dichotomous one ($n = 32$) by collapsing the responses *never*, *seldom*, and *occasionally* into negative responses (indicating *unreliable ecological engagement*) and *often* and *always* into positive responses (indicating *reliable ecological engagement*). This measure is in line with previous scale calibrations (e.g., Kaiser & Wilson, 2004). The measure had to be taken because participants generally are found to be inconsistent in using the more diverse response alternatives (see Kaiser & Wilson, 2000). Furthermore, negative responses to negatively formulated items (i.e., unecological behaviors) were recoded as *ecological engagement* responses and vice versa. In all 50 items, *I cannot answer* was coded as a missing value.

Overall conservation behavior was assessed using the Rasch model (for model details, see Embretson & Reise, 2000). In the first assessment, based on the first data set, 6.8% of all behavior responses were missing. The behavior scale’s Rasch-based reliability coefficient was .77 (for details, see Wright & Masters, 1982). In the second assessment, based on the second data set, 8.6% of the answers to the 50 behavior items were missing. The scale’s reliability this time was .74.

In conclusion, the two scale calibrations of the two compound behavior measures are in line with prior research and corroborate the fact that all 50 actions fall within a single behavior domain, including the 48 behaviors used in the incompatible assessment of the four TPB components.

Evidence for the conservation performance measure's validity is provided elsewhere (for construct validity information, see Kaiser, 1998; for the self-report measure's link with tangible behavior, see Kaiser, Frick, & Stoll-Kleemann, 2001; for ecological validity information, see Kaiser, Doka, Hofstetter, & Ranney, 2003).

Statistical Analysis

Because of the statistical identifiability of the factor loadings, each of the four TPB components requires at least two indicators (cf. McDonald & Ho, 2002). For the specific models, we employed two items as these two indicators—distinct in their question-and-response format—each referring to a single conservation behavior. For the six compatibly measured models, we used the same six behaviors with each of the four TPB concepts twice. For the six incompatibly measured models, the two indicators relate to two distinct conservation behaviors for each of the four concepts. Thus, for each of the six incompatibly measured models, we surveyed eight separate conservation behaviors; two (distinct in their response formats) for each of the four TPB components (see Appendix).

For the general models, we averaged the six items that we had used with each of two response formats to assess the TPB concepts to form one of two parcels. For the compatibly measured models, these two parcels (i.e., the two indicators of a concept) twice represent the same six behaviors for each of the four TPB concepts. For the incompatibly measured models, the two parcels contain six unique behaviors each. Thus, for the incompatibly measured general model, we explored 48 conservation behaviors (see Appendix). Note that for endeavors like the present one, when the primary goal is theory testing and when items represent well-established, but not highly internally

Table 1

Means, Correlations, and Residuals of the Scales Used with the Incompatibly Measured General TPB Model

Subscale	M	SD	N	1	2	3	4	5	6	7	8	9
1. Att1	3.85	0.61	784	.00	.00	.03	-.03	-.01	-.03	.02	-.04	.02
2. Att2	3.34	0.70	779	.38	.00	.04	-.01	.04	.05	.00	-.01	-.03
3. SN1	3.65	0.85	781	.34	.26	.00	.00	-.03	-.01	.04	.00	-.05
4. SN2	3.22	0.65	777	.29	.21	.42	.00	.01	.05	-.02	.05	.01
5. PBC1	3.46	0.69	780	.43	.34	.23	.27	.00	.00	-.03	.01	.01
6. PBC2	3.71	0.69	781	.32	.29	.19	.26	.44	.00	-.04	.05	.00
7. Int1	3.83	0.58	780	.50	.33	.34	.29	.42	.32	.00	.02	.00
8. Int2	3.57	0.72	783	.42	.30	.28	.34	.43	.39	.48	.00	-.01
9. GEB	-0.09	0.81	787	.58	.36	.29	.36	.53	.42	.57	.52	.00

Note. Figures below the diagonal represent Pearson correlation coefficients, while figures above and including the diagonal present the residual values of the model in Figure 1. TPB = theory of planned behavior; GEB = General Ecological Behavior scale (Kaiser, 1998; conservation behavior measure); Att = attitude toward behavior; SN = subjective norms; PBC = perceived behavioral control; Int = behavior intention. Each of the four TPB concepts is assessed with two indicators (e.g., Att1, Att2).

consistent concept indicators, parceling is a methodologically reasonable and acceptable strategy (cf. Little, Cunningham, Shahar, & Widaman, 2002).

All structural equation models were assessed using the maximum likelihood method. They were all tested confirmatorily, without allowing for any model modification (cf. MacCallum, Roznowski, & Necowitz, 1992). Covariance matrices were used as input matrices. With the two general models, correcting for measurement error attenuation by constraining the behavior scales' reliabilities (and the corresponding error variances) to the previously established values from the Rasch scale calibrations was only possible by using the correlation matrix as the input matrix (see Table 1). Note that except for the proportions of explained behavior variances and the corresponding beta weights of intention, the different input matrices only had trivial effects on the numerical

solutions.

Besides the chi-square statistic, we present three widely used goodness-of-fit indices: comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square of residual (SRMR). Following traditional recommendations (e.g., Bentler & Bonett, 1980; Browne & Cudeck, 1993), cutoff values around .90 for CFI, and below .08 for RMSEA and SRMR are adopted as indicators of a reasonable fit between model-implied and observed data.

Results

Our findings are reported in two sections. First, the six specific TPB model tests are presented: compatibly measured and incompatibly measured ones. Second, results for the compatibly and incompatibly measured general TPB models are detailed.

Specific TPB Models

Without model modification, confirmatory testing of the six compatibly measured specific models—one for each of six behaviors (recycle glass, recycle paper, no car use downtown, pointing out others, organic foods, and environmental organization membership)—revealed the TPB to be an empirically acceptable model. All six specific model tests provide results that are consistent with previous findings (e.g., Madden et al., 1992).

Note that the chi-square statistic is affected by sample size, which is quite large ($N = 607$). Not surprisingly, then, the chi-square statistic was significant in all six model tests (see Table 2; top part). Note also that the 18 model fit indicators, which are relatively independent of sample size (i.e., CFI, RMSEA, SRMR), generally suggest an impressive overlap and acceptable discrepancies between the observed and model-implied figures (i.e., all CFI values exceed the .90 threshold, and all RMSEA and SRMR values are below .08). Overall, the fit statistics of all six specific models support the TPB remarkably. Table 3 (top) summarizes the five paths' coefficients described by the TPB across all six specific models. Again, these figures corroborate

previous research (e.g., Ajzen, 1991).

The six incompatibly measured specific models were also tested without allowing for any model modification. Again, the sample size is quite large ($N = 787$). Not surprisingly, in five of the six model tests, the chi-square statistic was significant (see bottom part of Table 2). Two models did not converge, even after 1,000 iterations. A third model reveals a rather poor fit between model-implied and observed data, even with the indicators, which are relatively independent of sample size (i.e., CFI value below .90 and RMSEA value at .10; see Table 2). The other three models yield overall model fit indicators that look acceptable (CFI values exceeding .90 and RMSEA and SRMR values below .08). Unanimously, though, we found at least one implausible estimate (e.g., negative error variances or Heywood case; Pearson correlation coefficients exceeding 1.00) in every incompatibly measured specific model. Examples of implausible values, such as beta weights that exceed 1.00 (which, in turn, often either imply Heywood cases or alternatively suppressor effects within structural equation models), can be found in Table 3 (bottom part).

Next to implausible values and despite some high figures in absolute terms, we also found either barely or inconsistently confirmed theoretical relationships in every incompatibly measured model (see bottom part of Table 3). Of the five theoretical paths in the six conducted model tests (i.e., 30 paths in total), only eight turned out to be statistically significant: One of these was implausible, and one was in an unexpected, reverse direction (i.e., negative relationship). Unlike the results for the compatibly measured models, where 27 of the 30 paths (90%) were significant and as predicted for the incompatible models, 24 of the 30 paths (80%) were nonsignificant, implausible, or reversed (see Table 3). With regard to the behavior variance, 50.2% on average was explained in the compatible models, compared to 3.2% in the incompatible ones. As predicted, all of the incompatibly measured specific models more or less failed to corroborate the TPB in its ability to explain the six self-reported conservation behaviors in question.

Table 2

Fit Statistics for the Specific TPB Models

	$\chi^2(20)$	<i>p</i>	SRMR	RMSEA	CFI
Compatibly measured models (<i>N</i> = 607)					
1. Recycle glass	84.13	<.001	.045	.073	.94
2. Recycle paper	32.62	<.037	.027	.032	.99
3. No car use downtown	42.28	<.003	.020	.043	1.00
4. Point out others	48.53	<.001	.020	.049	1.00
5. Organic foods	66.53	<.001	.031	.062	.99
6. Environmental organization membership	45.80	<.001	.028	.046	.99
Incompatibly measured models (<i>N</i> = 787)					
1. Recycle glass	183.84	<.001	.066	.100	.86
2. Recycle paper	35.42	<.018	.027	.031	.96
3. No car use downtown ^a	25.96	<.170	.024	.019	.97
4. Point out others	47.09	<.001	.030	.042	.97
5. Organic foods ^a	58.35	<.001	.039	.049	.93
6. Environmental organization membership	32.25	<.041	.027	.028	.97

Note. TPB = theory of planned behavior; SRMR = standardized root mean square of residual; RMSEA = root mean error of approximation; CFI = comparative fit index. ^aThese two models did not converge, even after 1,000 iterations. Raising the default convergence criterion (by a factor of 1,000) yielded tentative solutions.

Table 3

Standardized Multiple Regression Coefficients of the TPB Paths Across the Specific TPB Models

	Att → Int	SN → Int	PBC → Int	PBC → Beh	Int → Beh
Compatibly measured models (<i>N</i> = 607)					
1. Recycle glass	0.74***	-0.01	-0.05	0.18***	0.55***
2. Recycle paper	0.71***	0.21*	-0.14	0.15**	0.47***
3. No car use downtown	0.34***	0.17***	0.55***	0.17**	0.70***
4. Point out others	0.48***	0.31***	0.14*	0.28***	0.54***
5. Organic foods	0.60***	0.26**	0.15***	0.15***	0.70***
6. Environmental organization membership	0.31***	0.30***	0.27***	0.22***	0.47***
Incompatibly measured models (<i>N</i> = 787)					
1. Recycle glass	0.41*	0.17	0.72***	-0.52***	0.59***
2. Recycle paper	0.38	0.86*	-0.19	0.11	0.25***
3. No car use downtown ^a	0.07	0.00	0.93	2.51	-2.46
4. Point out others	0.40	0.10	0.21	0.48***	-0.18
5. Organic foods ^a	-0.44	0.48	1.12**	0.73	-0.40
6. Environmental organization membership	0.26	-0.16	0.73	0.09	0.05

Note. TPB = theory of planned behavior; Att = attitude toward behavior; SN = subjective norms; PBC = perceived behavioral control; Int = behavior intention; Beh = behavior.

^aThese two models did not converge even after 1,000 iterations. Raising the default convergence criterion (by a factor of 1,000)

yielded tentative solutions. * $p < .05$. ** $p < .01$. *** $p < .001$.

General TPB Models

We proceeded to test two general versions of the theory by collapsing the six models reported previously: one at a compatible level and a second one at an incompatible level. For the compatible model test, we created two parcels as indicators for each of the four TPB concepts. These eight parcels each relate to the same six behaviors. For the incompatible model test, we also created two parcels as indicators for each of the four TPB concepts. This time, though, each of the eight parcels relates to a set of six unique behaviors (see Appendix).

As expected, a compatibly measured general TPB model can be accepted from an empirical point of view $\chi^2(21, N = 607) = 54.86, p < .001$ (SRMR = .026; RMSEA = .052; CFI = .99). Attitude (Att; $\beta = .57$), subjective norms (SN; $\beta = .25$), and PBC ($\beta = .19$) jointly explained 71% of the variance in behavior intention. They themselves correlated with one another considerably: $r_{(\text{Att-SN})} = .52$; $r_{(\text{SN-PBC})} = .41$; $r_{(\text{Att-PBC})} = .51$. A single factor, behavior intention ($\beta = .70$), determined 49% of the variance in a person's conservation behavior.

Surprisingly, an incompatibly measured general TPB model can also be accepted from an empirical point of view $\chi^2(21, N = 787) = 57.65, p < .001$ (SRMR = .024; RMSEA = .047; CFI = .99; see Figure 1). For the incompatible model, the determinants of attitude ($\beta = .60$), subjective norms ($\beta = .08; p > .05$), and PBC ($\beta = .35$), explained 91% of the variance in behavior intention. They themselves correlated with one another considerably: $r_{(\text{Att-SN})} = .66$; $r_{(\text{SN-PBC})} = .55$; $r_{(\text{Att-PBC})} = .79$. A single factor, behavior intention ($\beta = .94$), controlled 88% of the variance in a person's conservation behavior.

Measurement error arising from unreliable aspects of the behavior scale ($\sigma_e^2 = .26$) attenuates the influence of intention on behavior. Without correcting for measurement error attenuation by assuming a perfect behavior measure ($\sigma_e^2 = .00$), the proportion of explained variance of conservation behavior dropped to 65%, reducing intention's influence to a nevertheless considerable beta of .81. Despite this modification, the incompatibly mea-

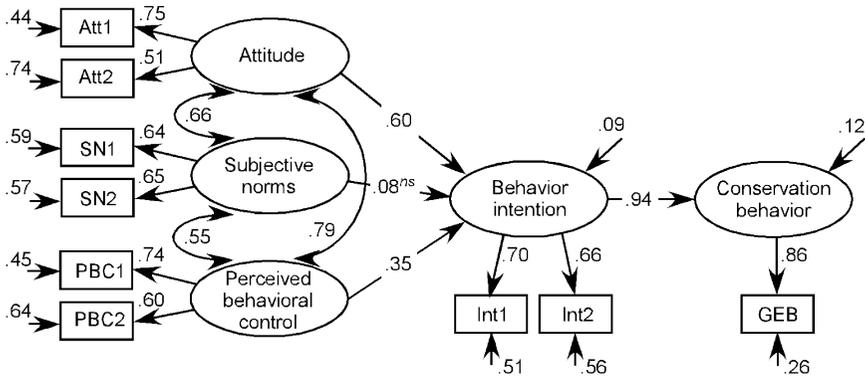


Figure 1. The incompatibly measured general theory of planned behavior ($N = 787$). Conservation behavior is measured with the General Ecological Behavior scale (Kaiser, 1998). The two indicators of attitude, subjective norms, perceived behavioral control, and behavior intention (e.g., Att1, Att2) represent two parcels composed of six distinct behavior items each (see Appendix). Because the relations between constructs are directed, arrows indicate such relations. Beta coefficients (i.e., standardized multiple regression coefficients) represent their strength. *ns* = nonsignificant relationship. Two-headed arrows are used when two concepts covary. These covariations are quantified with Pearson correlation coefficients. Arrows without origin indicate proportions of error and unexplained variances. The error variance ($\sigma_e^2 = .26$) and, accordingly, the reliability index (i.e., the square root of the reliability coefficient .74) of the conservation behavior measure ($\beta = .86$) are not estimated by the structural equation approach. Rather, they are derived from the Rasch model test.

sured general model's overall fit statistics remained unaffected: $\chi^2(21, N = 787) = 57.65, p < .001$ (SRMR = .024; RMSEA = .047; CFI = .99).

The quality of a model becomes particularly obvious in the accuracy of its prediction of the observed correlation matrix (McDonald & Ho, 2002). In other words, acceptable models should produce small discrepancies (i.e., residuals) between observed and model-implied correlations. This is the case with the incompatibly measured general TPB model for the overall (i.e., SRMR = .024), as well as the individual comparison of residuals. The latter figures vary, at most, by an amount of $\pm .05$ (see Table 1). Still, even if the overall fit of a model appears acceptable, the theoretically substantial paths of the model can prove to be poorly specified (McDonald & Ho, 2002). To assess

the quality of the theoretically meaningful relations, it is necessary to split the overall fit statistics (a) the measurement model fit statistics; and (b) the fit statistics into its two composites corresponding, for the entire model (see Figure 1) to the measurement model and the theoretically substantial model.³

Table 4

Separate Fit Statistics for the Overall, Measurement, and Theoretically Substantial Incompatibly Measured General TPB Model

Model	χ^2	<i>df</i>	<i>p</i>	RMSEA	<i>d</i>
Overall	57.65	21	<.001	.047	.047
Measurement	51.72	18	<.001	.049	.043
Substantial	5.93	3	<.115	.037	.004

Note. TPB = theory of planned behavior; RMSEA = root mean error of approximation; *d* = noncentrality index.

After subtracting the contribution of the measurement model from the fit statistics of the overall model, Table 4 reveals that both the measurement and the theoretically substantial model fit the data acceptably. From an empirical point of view, we can conclude that the theoretically meaningful part of the incompatibly measured TPB seems to be accurately specified and properly identifies the relations among its concepts.

Since we used measures with partly identical content—12 behaviors overlap—on the predictor (i.e., intention) as well as on the criterion (i.e., behavior) side, any strong empirical relation between intention and behavior could still be trivial and

³ The chi-square residual of the overall minus the measurement model ($\Delta\chi^2$) represents the chi-square fit statistic for the theoretical substantial model, which can be tested for its statistical significance (cf. McDonald & Ho, 2002). Note that next to the chi-square values, the degrees of freedom are additive as well. For fit indicators that are relatively independent of sample size (e.g., RMSEA), the noncentrality index (*d*) must be calculated (for more details and formulas, see McDonald & Ho, 2002).

method-induced (for an elaborated argument, see Greve, 2001). While establishing a fully incompatible model, we dropped the 12 behaviors that we had used to measure intention from the overall performance measure, ending up with a 38-item conservation behavior scale. Neither the model's general fit statistics— $\chi^2(21, N = 787) = 63.27, p < .001$ (SRMR = .026; RMSEA = .051; CFI = .99)—nor the proportion of explained behavior variance deteriorated dramatically. Still, 76% of the original 88% behavior variance was explained by the single predictor of behavior intention ($\beta = .87$).

Discussion

In the present study, we found that the TPB depends on the common implementation of the compatibility principle when applied specifically to a single behavior. While the compatibly measured models closely match previous findings (e.g., Ajzen, 1991; Armitage & Conner, 2001), the incompatibly measured specific models yielded both a comparatively poor explanatory power (i.e., average proportion of explained behavior variances of 3% vs. 50%) and feeble support for the theory-implied relationships (see Table 3). Note that the breakdown of the incompatible models occurred even though the behaviors were fully compatible with respect to their target, as they all represent conservation behaviors (see Kaiser & Wilson, 2004). Strictly speaking, our so-called incompatible models mainly violate the compatibility principle regarding the particular actions involved and the context of their performance, but not with regard to the general target at which they are directed (cf. Ajzen & Fishbein, 2005).

Theoretically, the lack of common method variance in the incompatible items could have diluted the theory-implied relationships, as well as exaggerated them in the compatible items by adding shared method variance (see Campbell & O'Connell, 1982). Based on the work of Campbell and O'Connell, it seems to us that an artificial increase in the strength of the relationships is the most likely consequence of shared method variance for our case. Thus, we have reason to believe that, in its common specific realization, the explanatory

power of the TPB is inflated because of common method variance. As a result, specific models—explaining single actions— can only contribute to an ambiguous understanding of the underlying relationships and, thus, they cannot conclusively reveal the origins of the behavior in question. Such a view is consistent with that of early critics of the theory (e.g., Dawes & Smith, 1985).

However, our data do not suggest that the TPB is invalid, particularly because our results reveal a different picture with the two general versions of the TPB, which are in line with prior tests (e.g., Kaiser & Gutscher, 2003; Kaiser et al., 2005). The original determinants—attitude, subjective norms, and PBC—accounted for 71% and 91% of people’s intentions in the compatible and incompatible versions of the model, respectively. Intention, in turn, explained 49% and a stunning 88% of the variance in conservation behavior (cf. Figure 1). The results also show considerable overlap between attitude, subjective norms, and PBC, with correlations between .41 and .79 being slightly higher in the incompatible case. Despite the oblique concept space, the three determinants of intention still represent distinguishable concepts (as the correlations are distinct from 1.00).

It should also be noted that the close relationship between intention and behavior cannot be attributed to method artifacts of two overly similar or even identical measures (cf. Greve, 2001). For one, they are constructed differently. While behavior is measured on a Rasch scale, intention, by contrast, is grounded in factor scores. For another, the model’s explanatory

power remains high, even if the item overlap of the predictor (i.e., intention) and the criterion (i.e., behavior) is effectively brought to 0. For that, we dropped the six (compatible case) or 12 (incompatible case) behaviors that we had used to assess intention from the behavior measure (cf. Appendix), ending up with either a 44-item or a 38-item conservation behavior scale. With these two modified outcome measures, intention still determined 43% and 76% of behavior variance in the compatible and incompatible models, respectively.

Since our research is based on some nonstandard methods, three possible shortcomings are worth mentioning. First, with regard to conservation behavior, social desirability could play a role. However, we found previously that the responses of mature adults, like the ones in our study, proved unbiased in their readiness to respond in ways they may have thought we wanted them to (see Kaiser, 1998). Second, self-reported practices from the current behavior scale reveal satisfactory correspondence to tangible behavior ($k = .78$; see Kaiser et al., 2001). Third, the parceling procedure, which we used with the general versions of the TPB, only marginally affects findings of this kind of model test (see Kaiser et al., 2005). Moreover, parceling can be expected to result in more replicable findings on the conceptual level, as the procedure generally enhances the internal consistency of the concept measures (Little et al., 2002).

Improving the reliability of the attitude and behavior measures and, ultimately, intensifying the attitude–behavior link, rests on the principle of aggregation (cf. Ajzen & Fishbein, 2005). Unfortunately, aggregation is not possible with the traditional version of the TPB because PBC necessarily fails as a proxy for overall actual control (Kaiser & Gutscher, 2003). Hence, a general version of the TPB depends on a behavior criterion that directly incorporates the control-relevant influences external to a person by applying the Rasch model to the measurement of people’s behavior (e.g., see Kaiser, 1998). With this modification in the criterion and except for omitting the direct influence of PBC on behavior, the general version of the TPB otherwise remains unchanged.

With aggregation and by freeing the TPB from literal compatibility and, thus, from method-related artificial influences, the theory can predict behavior with an accuracy of 75–90%. Thus, TPB not only corroborates its sufficiency, but also its psychological significance, at least with respect to classes of behaviors (cf. Ajzen & Fishbein, 2005). As a result, empirical research based on general versions of the TPB seems capable of contributing to an unambiguous understanding and, hence, to the solution of societal problems, as it effectively reveals the psychological origins of people's behavior. The price social psychology must pay, in return, is a more traitlike conception of people's attitudes. This ties in with more classical definitions of a relatively enduring organization of beliefs, feelings, and thoughts (e.g., Rokeach, 1968; cf. Ajzen, 1988), from which our discipline has moved away (cf. Schwarz, 2000).

References

- Ajzen, I. (1988). *Attitudes, personality, and behavior*. Chicago: Dorsey.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211.
- Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, 52, 27–58.
- Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84, 888–918.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173–221). Mahwah, NJ: Lawrence Erlbaum.
- Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*, 22, 453–474.
- Ajzen, I., & Sexton, J. (1999). Depth of processing, belief congruence, and attitude–behavior correspondence. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 117–140). New York: Guilford.
- Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behavior: A meta-analytic review. *British Journal of Social Psychology*, 40, 471–499.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588–606.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Campbell, D. T., & O’Connell, E. J. (1982). Methods as diluting trait relationships rather than adding irrelevant systematic variance. In D. Brinberg & L. Kidder (Eds.), *New*

directions for methodology of social and behavioral science: Forms of validity in research (pp. 93–111). San Francisco: Jossey-Bass.

- Dawes, R. M., & Smith, T. L. (1985). Attitude and opinion measurement. In Lindsey & E. Aronsson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 1, pp. 509–566). New York: Random House.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Eid, M. (2000). A multitrait–multimethod model with minimal assumptions. *Psychometrika*, *65*, 241–261.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, *51*, 360–392.
- Fishbein, M., & Ajzen, I. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review*, *81*, 59–74.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Greve, W. (2001). Traps and gaps in action explanation: Theoretical problems of a psychology of human action. *Psychological Review*, *108*, 435–451.
- Kaiser, F. G. (1998). A general measure of ecological behavior. *Journal of Applied Social Psychology*, *28*, 395–422.
- Kaiser, F. G., Doka, G., Hofstetter, P., & Ranney, M. A. (2003). Ecological behavior and its environmental consequences: A life cycle assessment of a self-report measure. *Journal of Environmental Psychology*, *23*, 11–20.
- Kaiser, F. G., Frick, J., & Stoll-Kleemann, S. (2001). Zur Angemessenheit selbstberichteten Verhaltens: Eine Validitätsuntersuchung der Skala Allgemeinen Ökologischen Verhaltens [Accuracy of self-reports: Validating the General Ecological Behavior scale]. *Diagnostica*, *47*, 88–95.
- Kaiser, F. G., & Gutscher, H. (2003). The proposition of a general version of the theory of planned behavior: Predicting

- ecological behavior. *Journal of Applied Social Psychology*, *33*, 586–603.
- Kaiser, F. G., Hübner, G., & Bogner, F. X. (2005). Contrasting the theory of planned behavior with the value–belief–norm model in explaining conservation behavior. *Journal of Applied Social Psychology*, *35*, 2150–2170.
- Kaiser, F. G., & Scheuthle, H. (2003). Two challenges to a moral extension of the theory of planned behavior: Moral norms and just-world beliefs in conservationism. *Personality and Individual Differences*, *35*, 1033–1048.
- Kaiser, F. G., & Wilson, M. (2000). Assessing people’s general ecological behavior: A cross-cultural measure. *Journal of Applied Social Psychology*, *30*, 952–978.
- Kaiser, F. G., & Wilson, M. (2004). Goal-directed conservation behavior: The specific composition of a general performance. *Personality and Individual Differences*, *36*, 1531–1544.
- Kirkpatrick, L. A. (1997). Effects of multiple determinacy and measurement error on trait–behavior and behavior–behavior relations: An integrated conceptual model. *Personality and Social Psychology Bulletin*, *23*, 199–209.
- Kothandapani, V. (1971). Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. *Journal of Personality and Social Psychology*, *19*, 321–333.
- Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, *86*, 114–121.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*, 151–173.
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490–504.
- Madden, T. J., Ellen, P. S., & Ajzen, I. (1992). A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and Social Psychology Bulletin*, *18*, 3–9.

- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analysis. *Psychological Methods*, 7, 64–82.
- Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, 100, 398–407.
- Oskamp, S., & Schultz, P. W. (1998). *Applied social psychology* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Rokeach, M. (1968). *Beliefs, attitudes, and values: A theory of organization and change*. San Francisco: Jossey-Bass.
- Schwarz, N. (1999). Self-reports: How questions shape the answers. *American Psychologist*, 54, 93–105.
- Schwarz, N. (2000). Social judgment and attitudes: Warmer, more social, and less conscious. *European Journal of Social Psychology*, 30, 149–176.
- Sheeran, P., & Orbell, S. (1998). Do intentions predict condom use? Meta-analysis and examination of six moderator variables. *British Journal of Social Psychology*, 37, 231–250.
- Weigel, R. H., & Newman, L. S. (1976). Increasing attitude–behavior correspondence by broadening the scope of the behavioral measure. *Journal of Personality and Social Psychology*, 33, 793–802.
- Wright, B., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA.

Appendix

Behaviors Used in the Compatible and Incompatible Measurement of the Four Theory of Planned Behavior Concepts

		Attitude		Subjective norms		Perceived control		Intention	
		<i>good–bad</i>	<i>appropriate–inappropriate</i>	<i>likely–unlikely</i>	<i>agree–disagree</i>	<i>easy–difficult</i>	<i>simple–complicated</i>	<i>likely–unlikely</i>	<i>decided–undecided</i>
Compatible measurement (first data collection)	1	I bring empty bottles to a recycling bin.	I bring empty bottles to a recycling bin.	I bring empty bottles to a recycling bin.	I bring empty bottles to a recycling bin.	I bring empty bottles to a recycling bin.	I bring empty bottles to a recycling bin.	I bring empty bottles to a recycling bin.	I bring empty bottles to a recycling bin.
	2	I collect and recycle used paper.	I collect and recycle used paper.	I collect and recycle used paper.	I collect and recycle used paper.	I collect and recycle used paper.	I collect and recycle used paper.	I collect and recycle used paper.	I collect and recycle used paper.
	3	I refrain from driving my car in or into the city.	I refrain from driving my car in or into the city.	I refrain from driving my car in or into the city.	I refrain from driving my car in or into the city.	I refrain from driving my car in or into the city.	I refrain from driving my car in or into the city.	I refrain from driving my car in or into the city.	I refrain from driving my car in or into the city.
	4	I have pointed out unecological behavior to someone.	I have pointed out unecological behavior to someone.	I have pointed out unecological behavior to someone.	I have pointed out unecological behavior to someone.	I have pointed out unecological behavior to someone.	I have pointed out unecological behavior to someone.	I have pointed out unecological behavior to someone.	I have pointed out unecological behavior to someone.
	5	I buy organically grown produce.	I buy organically grown produce.	I buy organically grown produce.	I buy organically grown produce.	I buy organically grown produce.	I buy organically grown produce.	I buy organically grown produce.	I buy organically grown produce.
	6	I am a member of an environmental organization.	I am a member of an environmental organization.	I am a member of an environmental organization.	I am a member of an environmental organization.	I am a member of an environmental organization.	I am a member of an environmental organization.	I am a member of an environmental organization.	I am a member of an environmental organization.
Incompatible measurement (second data collection)	1	I have pointed out unecological behavior to someone. (4)	At red traffic lights, I turn off the engine.	I own a fuel-efficient automobile (more than 33 miles per gallon).	I buy organically grown produce. (5)	In nearby areas (around 30 km; around 20 miles), I use public transportation or ride a bike.	I ride a bicycle or take public transportation to work or school.	I talk with friends about problems related to the environment.	I refrain from driving my car in or into the city. (3)
	2	I refrain from using a chemical air freshener in my bathroom.	I wash dirty clothes without prewashing.	I refrain from buying convenience foods.	I refrain from buying bleached and colored toilet paper.	In the winter, I refrain from keeping the heat on so that I do not have to wear a sweater.	In winter, I turn down the heat when I leave my apartment for more than 4 hours.	I buy products in refillable packages.	For longer journeys (more than 6 hours), I refrain from taking an airplane.

Appendix
Continued

	Attitude		Subjective norms		Perceived control		Intention	
	<i>good–bad</i>	<i>appropriate–inappropriate</i>	<i>likely–unlikely</i>	<i>agree–disagree</i>	<i>easy–difficult</i>	<i>simple–complicated</i>	<i>likely–unlikely</i>	<i>decided– undecided</i>
3	I bring empty bottles to a recycling bin. (1)	In hotels, I refrain from having the towels changed daily.	I reuse my shopping bags.	I wait until I have a full load before doing my laundry.	I prefer to shower rather than to take a bath.	I collect and recycle used paper. (2)	After a picnic, leave the place as clean as it was originally.	I refrain from putting dead batteries in the garbage.
4	I refrain from using a clothes dryer.	I am a member of an environmental organization. (6)	I read about environmental issues.	I have looked into the pros and cons of having a private source of solar power.	I contribute financially to environmental organizations.	I buy milk in returnable bottles.	I refrain from owning a car.	I use renewable energy sources.
5	I buy seasonal produce.	I refrain from using an oven cleaning spray to clean my oven.	I own energy-efficient household devices.	I drive in such a way as to keep my fuel consumption as low as possible.	I refrain from killing insects with a chemical insecticide.	I turn off the engine while waiting in front of a railroad crossing or in a traffic jam.	I refrain from buying beverages in cans.	After meals, I refrain from disposing of leftovers in the toilet.
6	drive on freeways at speeds under 100 kph (= 62.5 mph).	I have purchased solar panels to produce energy.	I refrain from using fabric softener with my laundry.	If I am offered a plastic bag in a store, I refrain from taking it.	I refrain from driving to where I want to start my hikes.	I boycott companies with an unecological background.	In the winter, I refrain from leaving the windows open for long periods of time to let in fresh air.	I buy furniture made from domestically grown wood.
Indicator	1	2	1	2	1	2	1	2

Note. Each of the six rows within the compatible and the incompatible approaches stands for one of the six specific models tested. The compound of all six behaviors per response format per concept represents one indicator in one of the general model tests. The shaded cells give the behavioral criteria for the incompatible Models 1 through 6.