Psychological Review
2008, Vol. 115, No. 2, 447–462

0033-295X/08/$12.00 DOI: 10.1037/0033-295X.115.2.447

Potential Performance Theory (PPT): A General Theory of Task Performance Applied to Morality

David Trafimow and Stephen Rice
New Mexico State University

People can use a variety of different strategies to perform tasks and these strategies all have two characteristics in common. First, they can be evaluated in comparison with either an absolute or a relative standard. Second, they can be used at varying levels of consistency. In the present article, the authors develop a general theory of task performance called *potential performance theory* (PPT) that distinguishes between observed scores and true scores that are corrected for inconsistency (i.e., potential scores). In addition, they argue that any kind of improvement to task performance, whatever it may be, works by influencing either task strategies, which comprise all nonrandom components that are relevant to the task, or the consistency with which strategies are used. In the current study, PPT is used to demonstrate how task strategies and the consistencies with which they are used impact actual performance in the domain of morality. These conclusions are extended to other domains of task performance.

Keywords: performance, true, score, morality, theory

Many human activities can be conceptualized as involving task performance, although perhaps the term *task performance* is not always used to characterize the behaviors involved in these activities. For tasks such as visual search, operating a vehicle, or performing surgery, task performance is widely used as an indicator of behavior. However, in other situations, one might not consider task performance the appropriate term, although there is no reason why it cannot be used in these situations as well. For example, when two people need to agree on a course of action (e.g., which movie to see, which restaurant to go to), it can be considered a failure if they cannot agree or if continual disagreement leads to stress or dissent. Repeated failures are likely to lead to the termination of the relationship. Thus, in this case, one might consider task performance to be a potential indicator of success or failure. One of our goals, then, is to emphasize and demonstrate that human activities that are not generally considered task performance behaviors can profitably be conceptualized as such.

A second goal pertains to the specificity with which theories predict findings. Trafimow (2003) demonstrated, quantitatively, the importance of proposing theories that make precise predictions (also see Roberts & Pashler, 2000). Yet, most task performance theories make only relative predictions across experimental conditions—that participants in one condition should perform better than participants in another condition—and not absolute predictions about individual behavior. For example, researchers might predict that voice recognition software reduces task interference relative to visual and/or manual control in computer design (e.g., Martin, 1989), but they may not be able to predict a priori exactly how big that difference is for each individual. There is no intention here to criticize previous task performance literature on the basis of a lack of specificity. Indeed, we believe that even vague predictions can nevertheless have value. But there can be little doubt that absolute prediction within conditions and, better yet, within individuals would be a further advance and would augment falsifiability.

For the purpose of this article, we have chosen to develop a general theory of task performance—*potential performance theory* (PPT)—in the domain of morality; thus, successful moral decision making is our measure of task performance. We chose this route for three reasons: (a) Morality is an issue that is important to all of humanity and transcends culture and geography; (b) by using a particular example with which to develop and outline our theory, we can avoid the typical confusion involved in presenting abstract mathematical concepts in absentia of specific examples; and (c) morality lends itself easily to making relative comparisons in absence of an absolute correct answer. Thus, our third goal is to develop, within the context of a nonobvious type of task performance behavior, a theory of task performance that includes both the goodness of task performance strategies and the consistencies with which they are used. Nevertheless, we stress that PPT is a general theory of task performance that is not limited to the moral domain.

Distinguishing Between Moral Conceptions and Moral Behaviors

From time immemorial, people have been interested not only in making specific moral decisions but also in the general principles, processes, or strategies by which those decisions should be made. Some proposals are that the strategies derive or should derive from a higher authority such as God, the law, the king, or others (e.g., Hobbes, 1651/1957); enlightened self-interest (e.g., Smith, 1759/

David Trafimow and Stephen Rice, Department of Psychology, New Mexico State University.

We thank Amy Wells and Krisstal Clayton for their help in collecting data.

Correspondence concerning this article should be addressed to David Trafimow, Department of Psychology, MSC 3452, New Mexico State University, P.O. Box 30001, Las Cruces, NM 88003-8001. E-mail: dtrafimo@nmsu.edu

1976); a feeling for the good of the community (e.g., Aristotle, trans. 2000); intuition (e.g., Hardin, 1982; Hume, 1740/1978; Thomson, 1989); utilitarianism (e.g., Mill, 1861/1979); duties (e.g., Kant, 1797/1991, Korsgaard, 1985); and so on. Social scientists have also entered the fray. The most famous example, perhaps, is Kohlberg, who proposed that people proceed through various stages of moral development (e.g., Kohlberg, 1981, 1984). Implicit in Kohlbergs theorizing is that some moral decision processes are more advanced than others. Not surprisingly, others have weighed in with alternative ideas (e.g., Gilligan, 1982; Gilligan & Attanucci, 1988).

Research in the domain of morality often includes a strong conceptual component that deals with issues such as evaluating the most defensible moral strategy, how moral strategies compare with each other, why moral strategies are important, how people develop moral strategies, how people use moral strategies, when they use moral strategies (or when they do not), and so on. Additionally, at least in psychological research pertaining to morality, there is an assumption that moral research has practical implications. Put bluntly, why should researchers study moral strategies if they do not influence human behaviors? Clearly, psychologists who study moral strategies make the assumption, often tacitly, that people's use of moral strategies does influence their behaviors. Even those who espouse some sort of intuitionism (e.g., Haidt, 2001; Hardin, 1982; Hume, 1740/1978; Thomson, 1989)—that people make moral decisions on the basis of an intuition—are assuming that a strategy is involved. The intuition presumably derives from some kind of mental process involving accessible concepts (Wyer & Srull, 1986), affects (Trafimow & Sheeran, 2004), or something else that may or may not have increased the likelihood of gene propagation during human evolutionary history (Johnston, 1999).

Because philosophers and psychologists who are interested in morality tend to pit different moral strategies against each other, either with a view to comparing how justifiable they are or to determining what people actually use, there is a tendency to not attend to the similarities among moral strategies. In much of normal living, different moral strategies actually suggest the same decisions, albeit for different reasons. For example, consider the decision about whether to kill a person in order to steal his money for a new Mercedes. A utilitarian might argue that this behavior increases the total harm more than the total good, and so the behavior should not be performed. A Jew, Christian, or Muslim might argue that performing the behavior contradicts God's commandments and so it should not be performed. An intuitionist might claim that the behavior simply feels wrong. Whatever the reasoning, the actual decision is the same: to not perform the behavior.

That different moral strategies often lead to similar conclusions suggests interesting possibilities, depending on whether one believes that there is an objectively correct moral strategy or whether one believes that there is not. From the former perspective, it seems that many incorrect moral strategies nevertheless lead to morally correct behaviors. From the latter perspective, it seems that many different moral strategies lead to similar behaviors. Either way, different strategies are likely to lead to a reasonable level of moral performance whether measured by agreement with some absolute standard or by agreement with other strategies.

But the moral strategy one uses is not the only factor; there is also the consistency with which the strategy is used. Inconsistent use of even an optimal strategy may well lead to suboptimal moral decision making. For example, a person may fully believe everything the Koran teaches; however, that same person may be inconsistent in following the guidelines and regulations laid out in the Koran. This inconsistency will naturally lead to suboptimal moral performance when compared with another person who consistently follows the moral teachings therein.

Consequently, our goal is to propose a theory that takes into account, among other things, moral strategies and the consistencies with which they are used. We expect to be able to use this theory to (a) mathematically simulate the relative importance of strategies and consistency under different conditions, (b) allow the prediction of how well a strategy will predict moral behavior at varying levels of consistency, (c) allow one to determine how much moral decision making can be improved by teaching people better strategies, and (d) allow one to determine how much moral decision making can be improved by teaching people to use their strategies more consistently.

Before deriving the theory, it is important to reiterate that there are *objectivists* who believe that there is at least one moral strategy independent of human mental processes and *relativists* who do not believe this. Consequently, from the former point of view, it makes sense to discuss whether behaviors are morally correct. From the latter perspective, the best that can be done is to discuss the percentage of agreement between people using different moral strategies. It is important that the equations to be presented can be made compatible with each viewpoint.

Converting From Tables to Correlations

Imagine that two people are presented with a set of moral decisions. Each person can decide one way or another way, which renders a 2 × 2 table with four possibilities: Persons A and B both choose Option 1, Person A chooses Option 1 and Person B chooses Option 2, Person A chooses Option 2 and Person B chooses Option 1, or Persons A and B both choose Option 2. Or, if there is only one person who is compared with some absolute standard, the result is still a 2 × 2 table where the person correctly chooses Option A, incorrectly chooses Option A, correctly chooses Option B, or incorrectly chooses Option B. Any 2 × 2 table can be converted into a correlation coefficient by using the phi coefficient equation presented below as Equation 1, which is a special case of the more general equation for computing the Cramér coefficient (e.g., Siegel & Castellan, 1988).

$$r_\phi = \frac{|ad - bc|}{\sqrt{r_1 r_2 c_1 c_2}} = \frac{|ad - bc|}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}\,. \qquad (1)$$

To understand Equation 1, imagine a 2 × 2 table where the first, second, third, and fourth cells are designated by a, b, c, and d, respectively; the first and second row totals are represented by r_1 and r_2, respectively; and the first and second column totals are represented by c_1 and c_2, respectively. Note that either margin totals or cell frequencies can be used in the denominator because that fact will turn out to be important.

We can now make use of classical true score theory (for accessible reviews, see Allen & Yen, 1979; Cohen & Swerdlik, 1999;

Crocker & Algina, 1986).[1] It is well-known that correlations can be corrected for attenuation because of inconsistency and Equation 2 is how this is done. In the equation below, R is the true, corrected, or potential correlation coefficient as a function of the observed or actual correlation coefficient (r), the consistency of person x ($r_{xx'}$), and the consistency of person y ($r_{yy'}$). Hereafter, the terms *true*, *corrected*, and *potential* are used interchangeably and are represented by uppercase letters (e.g., R), whereas the terms *observed* and *actual* are used interchangeably and are represented by lowercase letters (e.g., r). Finally, because the correction formula is general, we dropped the phi subscript from Equation 1.

$$R = \frac{r}{\sqrt{r_{xx'} r_{yy'}}}. \quad (2)$$

Converting True Correlations to Tables

Once the observed correlation has been adjusted via Equation 2, there remains the problem of converting back from the true correlation to a true 2 × 2 table. This can be done in two ways. One way is to make standard tables where the cells are expressed as row and column proportions and where the row and column margin values are automatically set at unity. Standard tables have the advantages of facilitating comparisons across persons and across experimental conditions. An additional advantage and one that we make much use of is that they provide a vehicle for demonstrating general relationships between observed successes, consistency, and true successes. But they have the disadvantage of losing some of the information contained in the observed tables; in particular, the original margin frequencies are lost except in the unlikely case that all of them were equal to each other in the original 2 × 2 table. The second way is to create nonstandard tables that preserve the original margin frequencies but where comparison across individuals and experimental conditions is more difficult. Both of these methods are demonstrated.

Standard Tables

Rosenthal and his colleagues (Rosenthal & Rosnow, 1991; Rosenthal & Rubin, 1979, 1982) have demonstrated that correlations can be converted into standard 2 × 2 tables with the following equation where S is the proportion of successes. In Table 1, their method was used to create a standard table based on the idea that the true correlation between two people is .3. Note that any standard table can be represented by a single arbitrarily chosen number. Throughout the present article, our arbitrary choice is to use the proportion of successes (S) as a function of the true correlation (R), as exemplified by the italicized value in Table 1 and as expressed in Equation 3.

$$S = \frac{(R + 1)}{2}. \quad (3)$$

Because all of the margin values are set at unity, it follows that the proportion of failures (F) is expressed by Equation 4.

$$F = 1 - S. \quad (4)$$

As we said earlier, the use of standard tables fails to take into account the original margin frequencies unless it so happens that they are all equal to each other. In this special case, it is always possible to convert from the observed percentage of successes (s) into the true percentage of successes (S) via Equation 5 when one is concerned with relative morality or via Equation 6 when one is concerned with absolute morality. Appendix A gives the derivation for Equation 5.

$$S = \frac{(2s - 1 + \sqrt{r_{xx'} r_{yy'}})}{2\sqrt{r_{xx'} r_{yy'}}}. \quad (5)$$

Equation 5 solves for the percentage of moral successes, defined as the agreement between a person who uses Moral Strategy X and a person who uses Moral Strategy Y, given each person's level of consistency at using their respective moral strategies. Equation 5 can be adapted easily for objectivists. If one believes that there is an objective moral strategy in the universe that could, at least hypothetically, be used consistently, then success is defined as agreement with what a hypothetical perfect moral entity would do. In that case, $r_{yy'}$ is defined as unity because the hypothetical perfect moral entity is perfectly consistent, and Equation 5 reduces to Equation 6 below.

$$S = \frac{(2s - 1 + \sqrt{r_{xx'}})}{2\sqrt{r_{xx'}}}. \quad (6)$$

Equations 5 and 6 give expected moral performance under perfect consistency, with relativist or objectivist assumptions, respectively. However, they do not give expected performance under varying levels of consistency. Fortunately, it is possible to develop equations that will serve this purpose using the foregoing equations as a starting point. The approach will be to use Equation 5 or Equation 6 to obtain moral performance under perfect consistency and then to adjust for any desired level of inconsistency.

Following this approach, the first step is to rearrange the terms from Equation 5 so that we solve for s rather than S. But remember that S is calculated on the basis of an s that is presumed to be an actual value in the real world that can be estimated from obtained data. So the goal is not to calculate s but rather to calculate what it would be under any particular level of consistency a researcher wishes to assume for the people involved. To reflect this goal, we refer to *desired* consistencies and observed successes instead of

Table 1
Standard Table for a Correlation Coefficient of .3

Person 1 choice	Moral	Immoral	Margin proportion
Person 2 choice			
Moral	*65%*	35%	100%
Immoral	35%	65%	100%
Margin proportion	100%	100%	

Note. Italics denote the use of proportion of successes as a function of true correlation.

[1] In the interest of simplicity, PPT is presently based on the assumptions of classical true score theory. We believe that PPT eventually may be made more liberal by using generalizability theory (Cronbach, Gleser, & Rajaratnam, 1963), but this is a matter that will have to be addressed in the future.

actually obtained ones, and we substitute appropriate subscripts, as can be seen in Equation 7.

$$s_{\text{DES}} = S\sqrt{r_{\text{DES-X}} r_{\text{DES-Y}}} + .5 - .5\sqrt{r_{\text{DES-X}} r_{\text{DES-Y}}}. \quad (7)$$

In Equation 7, s_{DES} is the desired percentage of moral successes when $r_{\text{DES-X}}$ is the desired level of consistency for using Moral Strategy X and $r_{\text{DES-Y}}$ is the desired consistency of using Moral Strategy Y. For objectivists, we assume that $r_{\text{DES-Y}}$ equals unity and so Equation 7 reduces to Equation 8 below.

$$s_{\text{DES}} = S\sqrt{r_{\text{DES-X}}} + .5 - .5\sqrt{r_{\text{DES-X}}}. \quad (8)$$

Nonstandard Tables

Although the standard tables described in the foregoing section are useful for making comparisons and for demonstrating the relations between actual successes, consistency, and true successes (as will become clear in following sections), under most situations they cause one to lose information about the original margin frequencies. If the original margin frequencies are important, the standard tables can be misleading. Therefore, we recommend that the use of standard tables be restricted to general demonstrations and that nonstandard tables be used to analyze real data.

As we have seen, it is easy to convert true correlations to standard tables because the fact that all of the margin proportions equal unity renders each table with only one unknown variable; if one cell is known, the others are determined. Thus, it is not particularly difficult to understand the simple mathematics rendered in Appendix A. In contrast, there are more degrees of freedom in the case of nonstandard tables and so it is necessary to solve simultaneous equations to obtain the nonstandard true cell frequencies. Consequently, Appendix B presents the necessary mathematics for deriving Equations 9–12, giving the true frequencies of Cells A–D, respectively. Note that the actual correlation can be obtained by applying Equation 1 to the original 2 × 2 table and that it can be converted to a true correlation (R) via Equation 2. In addition, analogous to the chi-square test and Fisher's exact test, the margin frequencies are set to equal those in the original 2 × 2 table (e.g., $R_1 = r_1$, $R_2 - r_2$, $C_1 = c_1$, and $C_2 = c_2$). Consequently, R, R_1, R_2, C_1, and C_2 are all obtainable from the data; in short, they are known values.

$$A = \frac{R\sqrt{R_1 R_2 C_1 C_2} + C_1 R_1}{(R_1 + R_2)}. \quad (9)$$

$$B = \frac{R_1(R_1 + R_2) - (R\sqrt{R_1 R_2 C_1 C_2} + C_1 R_1)}{(R_1 + R_2)}. \quad (10)$$

$$C = \frac{C_1 R_2 - R\sqrt{R_1 R_2 C_1 C_2}}{(R_1 + R_2)}. \quad (11)$$

$$D = \frac{C_2(R_1 + R_2) - [R_1(R_1 + R_2) - (R\sqrt{R_1 R_2 C_1 C_2} + C_1 R_1)]}{(R_1 + R_2)}. \quad (12)$$

Equations 9–12 define the nonstandard true cell frequencies in terms of values that can be obtained from the data and from the foregoing equations. For ease of calculation, however, once A is determined by using Equation 9, the other true cell frequencies can be computed merely by subtraction from the appropriate margin frequencies. Thus,

$$B = R_1 - A,$$

$$C = C_1 - A, \text{ and}$$

$$D = C_2 - B \text{ or } D = R_2 - C.$$

Using PPT to Understand Relations Among Variables

For most of the general points we make, it is convenient to use standard tables and to focus on Equations 5–8, although we reiterate our warning that nonstandard tables should be used to analyze real data. This standard paradigm is used until an experiment is presented for illustrative purposes.

Actual Performance Underestimates True Moral Performance

Equations 5 and 6 give the true percentage of moral successes after adjusting actual performance for attenuation due to inconsistency. These equations indicate that as consistency decreases, the extent to which actual moral performance underestimates true moral performance increases. Figure 1 renders true moral performance as a function of actual performance and consistency from an absolutist perspective (Equation 6). In general, as actual performance increases, there is less to be gained by correcting for attenuation due to inconsistency. And at very high levels of actual moral performance, a ceiling effect limits the effectiveness of such correction. But when actual performance is low, then there is a lot to be gained by correcting for attenuation due to inconsistency, particularly when consistency is also low. In this case, there seems to be a "sweet spot" where a small change in consistency (e.g., from .1 to .3 when $s = .6$) can make a large difference in the effectiveness of correcting for attenuation due to inconsistency.

Finally, there is an exception to the general rule that correction is more effective when actual performance is low. When actual moral performance is so low that it is random ($s = 50\%$), then correcting for attenuation due to inconsistency does not change anything. To understand the flat curve when s equals 50%, imagine that one's moral strategy is to flip a coin to determine each moral decision. Even if one used this strategy perfectly consistently, moral performance would remain at 50%. It is, of course, possible to use moral strategies that render moral performance at less than 50%. However, these were not graphed because they can be understood in the context of present figures merely by taking their complements. For example, if a researcher wished to gain a quick idea of what would happen if an actual level of moral performance of 40% was adjusted for attenuation due to inconsistency, the researcher could find the complement (60%) and find the appropriate curve in Figure 1. Then, for each of the values indicated for that curve, the complement could be imagined. The result, of course, is that one would imagine curves that go in the opposite direction of the curves in Figure 1. In general, when one uses a moral strategy that renders less than 50% actual successes, adjusting for inconsistency decreases true moral performance rather than increasing it. One would hope that few people would use such moral strategies, and they will not be considered further in the present article.

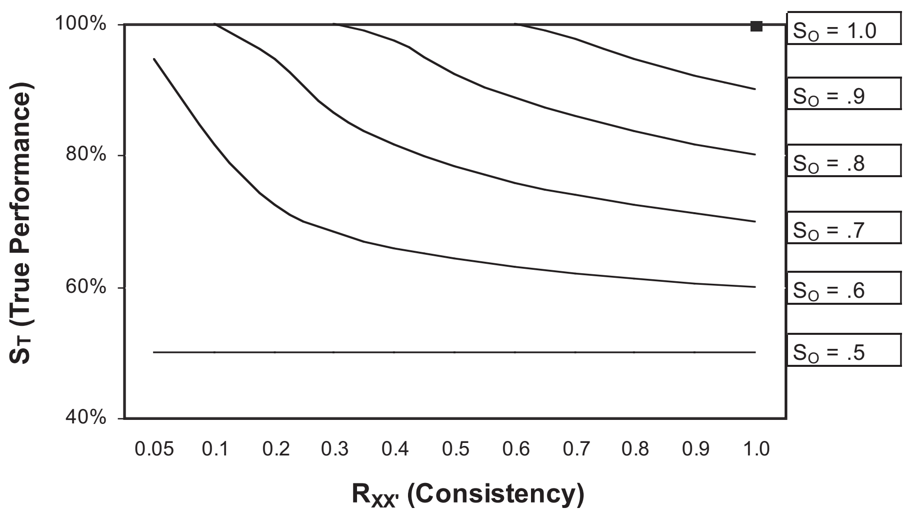

Figure 1. True moral performance (S_T) as a function of actual moral performance (S_O) and consistency ($R_{XX'}$) from an absolutist perspective based on Equation 6. The subscripts are added here to increase the clarity of the figures.

Figure 2 illustrates what happens from the relativist point of view, where it is necessary to adjust for the inconsistency of two people rather than one person (Equation 5). In Panel A, actual moral performance (s) was set at 51% rather than at 50% so that it would be possible to see at least a slight difference between the curves. Obviously, however, when actual moral performance is that low, there is little to be gained by adjusting for inconsistency. However, Panels B and C demonstrate a complicated interaction between actual moral performance and the consistencies of the two people. When both people are inconsistent, then adjusting has a dramatic effect, and this effect is even more dramatized as actual moral performance increases to 55% (Panel B) and 60% (Panel C). In summary, Figure 1 demonstrates the potential importance of adjusting for attenuation due to inconsistency when absolute morality is considered, and Figure 2 shows that such adjusting is even more important when relative morality is considered.

The Effects of Rule Goodness and Consistency on Actual Moral Performance

From an objectivist point of view, rule goodness is evaluated relative to an objective standard, whereas relativists would argue that rule goodness resides in agreement with others. These differing points of view are taken into account by Equation 7 and Equation 8. Let us start with Equation 8 and take on an objectivist position.

Figure 3 was created by using Equation 8, letting desired consistency vary between 0 and 1, and setting true moral performance (S) at .5, .6, .7, .8, .9, or 1.0. This figure shows a fascinating interaction between S and consistency. When S is low, then varying the level of consistency does not have much effect (e.g., the curve is flat when $S = .5$ and almost flat when $S = .6$). But when S is large, then the level of consistency has a major effect (e.g., the curve is extremely steep when $S = 1.0$). Figure 3 implies something interesting about moral training. If people are using moral strategies that, on being used consistently, would nevertheless fail to lead to an impressive frequency of correct moral decisions (e.g., 60%), then there is little point in training them to use these strategies consistently; it would be better to train them to use better strategies to begin with. But matters change dramatically for people who are already using good moral strategies, even if these strategies are not optimal. In that case, there is more to gain by training them to use their strategies consistently rather than trying to induce them to use even more optimal strategies. To see this, consider a person who uses a strategy with S equal to .9 and consistency equal to .3. Suppose that moral education is used to increase S to 1.0. The increase in frequency of moral successes would improve from 72% to 77%, an increase of only 5%. In contrast, suppose that moral education is used to increase consistency from .3 to 1.0. In that case, the increase in frequency of moral successes would improve from 72% to 90%, an increase of 18%.

Now consider the subjectivist position and Equation 7, as is illustrated in Figure 4. Here, there are three variables: true moral performance (S), the desired consistency of Moral Strategy X, and the desired consistency of Moral Strategy Y. Each panel represents a different level of the desired consistency of Moral Strategy Y. In Panel A, where the consistency of Y is set at .1, increasing the degree of consistency of X has limited effects, even when S takes on large values. But there is a complex interaction that can be seen as the consistency of Y increases to .4 (Panel B), .7 (Panel C), and 1.0 (Panel D). As Moral Strategy Y becomes more consistent, the interaction between the consistency of Moral Strategy X and true moral performance becomes more dramatic. In summary, Figure 3 demonstrates an important interaction between true moral performance and consistency on actual moral performance for absolute morality, and Figure 4 demonstrates that consistency plays an even more important role in this interaction for relative morality.

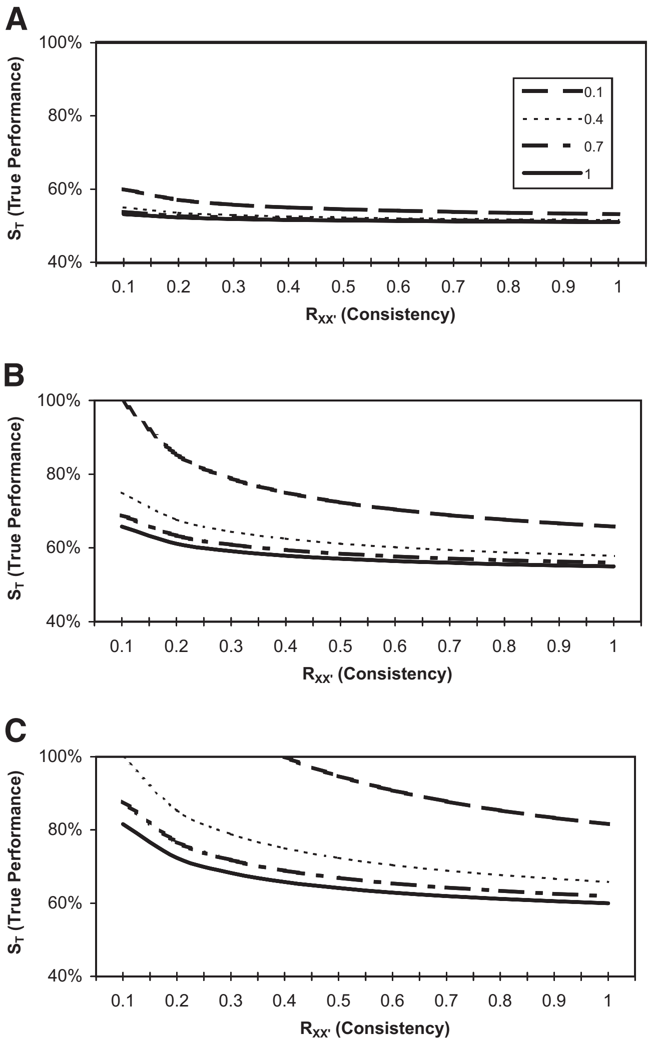

Figure 2. True moral performance (S_T) as a function of actual moral performance (S_O) and the consistencies of two people ($R_{XX'}$ and $R_{YY'}$) from a relativist perspective based on Equation 5. Actual moral performance (S_O) was set at 51%, 55%, and 60% in Panels A, B, and C, respectively.

Why Absolute Moral Performance Outpaces Relative Moral Performance

The mathematics of PPT suggests a counterintuitive conclusion about how absolute and relative morality relate to each other. Mathematically speaking, absolute morality is a special case of relative morality. Therefore, it is possible to compare and contrast them with each other. To see this, consider that only one consistency variable contributes to absolute moral performance (the consistency of Person X) but two consistency variables contribute to relative moral performance (the consistency of Person X and Person Y). Consequently, to the extent that people are not perfectly consistent, it implies that inconsistency should harm actual absolute moral performance less than it should harm actual relative moral performance. In addition, however, Equations 7 and 8 make quantification of this effect possible. All that is necessary to find the difference between relative and absolute performance (s_{DIFF} in Equation 13) is to subtract Equation 7 from Equation 8, which renders Equation 13 below.

$$s_{\mathrm{DIFF}} = S\sqrt{r_{\mathrm{DES\text{-}X}}} - .5\sqrt{r_{\mathrm{DES\text{-}X}}} - S\sqrt{r_{\mathrm{DES\text{-}X}}r_{\mathrm{DES\text{-}Y}}} + .5\sqrt{r_{\mathrm{DES\text{-}X}}r_{\mathrm{DES\text{-}Y}}}. \quad (13)$$

On the basis of Equation 13, Figure 5 illustrates what happens to the difference in actual moral performance when consistency ranges from 0 to 1 (for Figure 5, the consistencies of the two people are always kept equal to each other) and when true moral performance (*S*) is set at .5, .6, .7, .8, .9, or 1.0. As Figure 5 shows, when consistency is either extremely low or extremely high, the difference between relative and absolute performance tends toward zero. When consistency is extremely low, both absolute moral performance and relative moral performance tend toward randomness and consequently approach each other as well. And when consistency is extremely high, both absolute and relative moral performance tend toward true moral performance (and toward each other as well). However, when consistency is low but not too low, then the multiplication of inconsistencies decreases relative moral performance to a greater degree than absolute moral performance, and the difference between them is maximized.

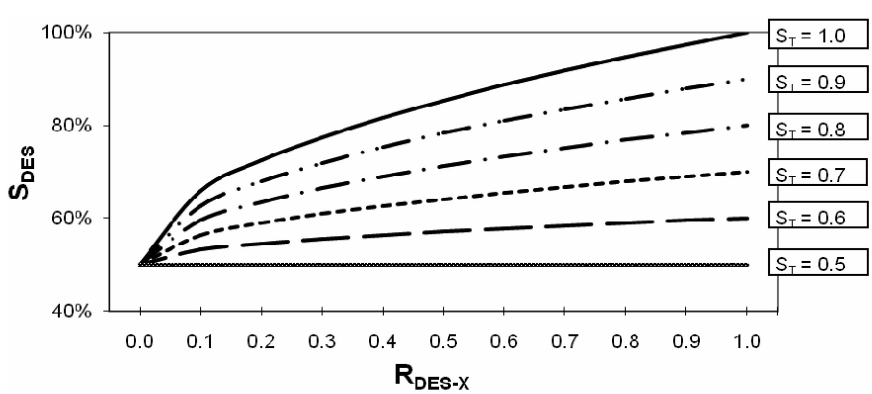

Figure 3. Desired percentage of successes (S_{DES}) as a function of true percentage of successes (S_T) and desired level of consistency ($R_{\mathrm{DES\text{-}X}}$) from an absolutist perspective based on Equation 8.

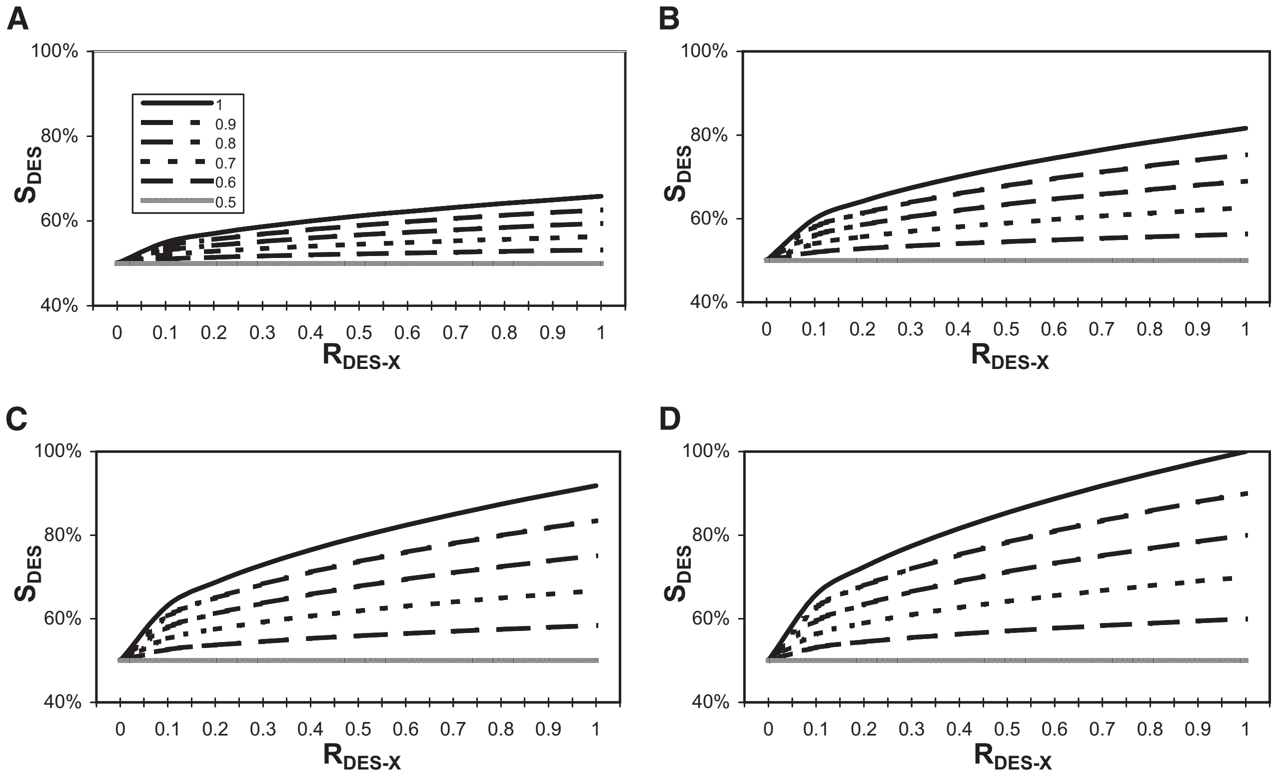

Figure 4. Desired percentage of successes (S_{DES}) as a function of true percentage of successes (S_T) and desired level of consistencies of two people ($R_{DES\text{-}X}$ and $R_{DES\text{-}Y}$) from a relativist perspective based on Equation 7. The desired level of consistency of the second person ($R_{DES\text{-}Y}$) was set at .1, .4, .7, and 1.0 in Panels A, B, C, and D, respectively.

Individual Differences in Morality

Many psychologists (e.g., Kohlberg, 1981, 1984) have defined morality in terms of a hierarchy of moral strategies. Thus, when researchers wish to place people along a dimension of morality, the people's placement depends on the strategies that the researchers infer from reports of the reasons for moral decisions. The present theory suggests three caveats for researchers that are accustomed to thinking in this way. First, people may vary in the consistency with which they use their moral strategies, which may well constitute an additional dimension of morality. That is, behaviors can be more or less moral because of people's strategies, or behaviors can be more or less moral because of the consistency with which people use their strategies. Second, from a practical perspective, if one insists on a single morality dimension, perhaps it should be the frequency of successful moral performances rather than either belief in moral strategies or consistency. Third, even if one is interested in the relation between beliefs in moral strategies and successfully making moral decisions (in either the objectivist or subjectivist paradigm), the present theory indicates that any kind of attempt to measure this relation is going to result in an underestimate of the true relation because of attenuation due to inconsistency.

To understand this last point, suppose a researcher performed a Kohlberg type of experiment where each participant was presented with a large set of moral choices and assigned a score indicating the extent to which that person's moral strategy resulted in good moral decisions, whether these decisions were considered in relation to some absolute standard (probably the experimenter's standard!) or in relation to someone else's use of some moral strategy. Equation 5 and Equation 6 indicate that the participant's observed score will be an underestimate of his or her true score and so the findings cannot be taken at face value.

Fortunately, PPT suggests an easy solution. Each participant can be presented with the large set of moral decisions at two separate time periods and the correlation between the two periods can be computed. The participant's absolute score is s and his or her correlation across the two time periods is $r_{xx'}$. Instantiating these values into Equation 6 allows an estimate to be made of the goodness of the person's true moral decision strategy, adjusted for inconsistency. For relativists who wish to test two people, s would be the percentage of times the two people actually agree and $r_{xx'}$ and $r_{yy'}$ are estimated by the correlations of each person across the two time periods. To obtain the best estimate of the true agreement between the two people, adjusted for inconsistency, these values would be instantiated into Equation 5.

How to Test and Use the Theory Empirically

Manipulate Consistency

Imagine that participants are asked to make a large number of moral decisions at Block 1, there is a delay, and then they are

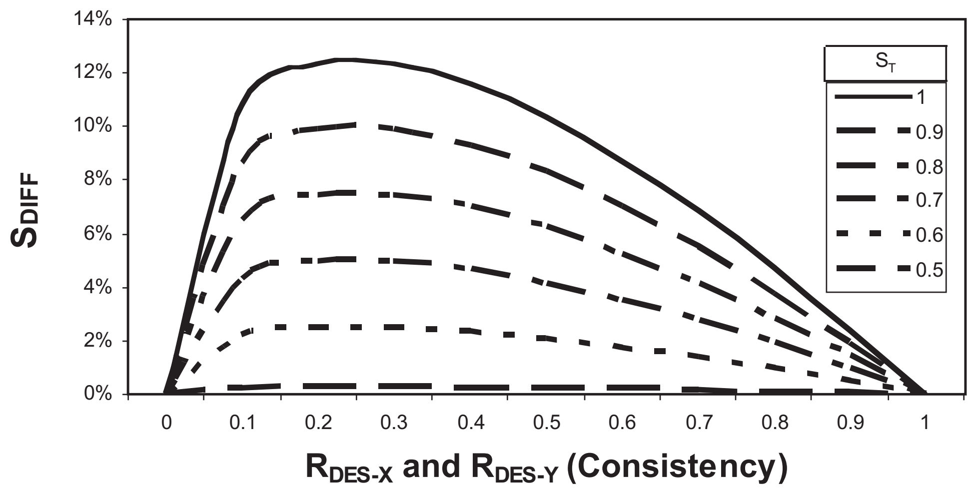

Figure 5. Difference between absolute and relative actual moral performance (S_{DIFF}) as a function of true moral performance (S_T) and desired consistency, based on Equation 13. The desired consistency of both people was always equal to each other (it is always the case that $R_{DES\text{-}X} = R_{DES\text{-}Y}$).

asked to make the moral decisions again at Block 2. By comparing each participant's responses between the two time periods, across all of the moral decisions, the experimenter can obtain a consistency coefficient ($r_{xx'}$ or $r_{yy'}$). In addition, it is easy to compute the actual percentage of moral successes by comparison with either an experimenter-generated absolute standard or other participants. Equation 6 (if an absolute standard is used) or Equation 5 (if a relativist standard is used) can be used to estimate the true percentage of successes that would be obtained with perfect consistency.

Now, suppose that the experimenter manipulated the participants' consistencies. Possibly, this could be done by instructing participants to be more consistent, by having participants practice making consistent moral decisions, by increasing the cognitive accessibility of their strategies, or by other means. Subsequent to the manipulation, the participants would again make moral decisions at two different time periods, and again consistency and actual performance would be computed. In the case of an absolute standard, Equation 8 would be used as follows. First, Equation 6 would be applied to the premanipulation data to obtain the true percentage of moral successes. This and the postmanipulation consistency would be instantiated into Equation 8, and the desired level of moral successes would be computed. This predicted percentage of moral successes would then be compared with the actual percentage of moral successes obtained in the postmanipulation phase of the experiment. To the extent that predicted percentages of moral successes and actually obtained levels of moral successes are close together, the theory would be supported. Of course, from a relativist perspective, a similar process would be followed but using Equations 5 and 7 rather than Equations 6 and 8.

Moral Training

Moral training plays an important role in society. Children go to religious school, there are civics courses in high schools, and there are courses on the philosophy of morality in universities. Even within the domain of psychology, researchers are required to read about or take courses on the ethics involved with using human participants. The distinction between the goodness of moral strategies and the consistency with which they are used implies a fascinating question about how moral training works. Does it work by inducing people to use better strategies, be more consistent in their use of moral strategies, or some degree of each? Fortunately, PPT can be used to answer this question. As in the foregoing subsection, the idea would be to obtain pretraining and posttraining measures of actual performance and consistency. From these, pretraining and posttraining percentages of true moral performance can be calculated. Thus, if consistency does not change as a result of moral training but true moral performance does, then any change in actual performance must be due to an increase in the goodness of people's moral strategies. But if true moral performance does not change from pretraining to posttraining, then any change in actual performance must be due to a change in consistency.

Of course, it is quite likely that moral training influences both the goodness of moral strategies and consistency, in which case it would be desirable to find out how much of each is changed and how these changes work together to increase actual moral performance. To find out how much training changes each of the variables, it is merely necessary to subtract pretraining from posttraining actual performance, consistency, and true performance. To find out how changes in strategy goodness and consistency work together to increase actual moral performance, a more complicated analysis is necessary, as is demonstrated in the following example.

Suppose that in the pretraining phase, a person's actual performance is 60%, her consistency coefficient is .3, and so her true moral performance, calculated by using Equation 6, is 68%. In the posttraining phase, these numbers are 80%, .8, and 84%, respectively. In that case, it can be shown that training increases actual performance from 60% to 80%, it increases consistency from .3 to .8, and it increases true moral performance from 68% to 84%. How important is the increase in true moral performance, the increase in consistency, and the interaction between them? Suppose that only true performance had increased but consistency remained constant. In that case, actual performance obtained from Equation 8 would be 69%, which is an increase from the original actual performance of 60%. Now consider what would have happened if only consistency had increased, in which case Equation 8 gives a value of 66%, which is an increase from the original value of 60%. Although it is certainly desirable to increase actual moral performance by 9 or 6 percentage points by increasing true moral performance or consistency, respectively, neither of these increases does justice to the interactive effect. Because the increase in actual performance in this example is 20 percentage points, it should be clear that the individual improvements in true moral performance and in consistency substantially underestimate the efficacy of the training; much of the efficacy is due to the interactive effect of gains in true moral performance and consistency.

There are additional ways of applying the theory to training. For example, suppose that a teacher is interested in training children to increase their actual moral performances but is not sure what to focus on in moral training: How much effort should be devoted to better strategies and how much effort should be devoted to increasing consistency? PPT provides an elegant way of answering this question. If the teacher measures children's actual moral performances and consistencies and uses the obtained values to obtain true values (using Equation 5 or Equation 6), the difference between the obtained and true values provides a strong clue about how to optimize training. If there is little difference between actual and true morality, then little is to be gained by increasing consistency, and the teacher would be better off focusing on moral strategies. However, if there is a large difference, then a great deal is to be gained by increasing consistency. This might be accomplished by so simple a means as repeated drills to increase the accessibility of whatever strategies the children are using. In addition, if the teacher wishes to be more sophisticated and optimize the training, he or she can use Equation 7 or Equation 8 to figure out the expected actual performance at various levels of true morality and consistency to determine exactly how much training effort to put into each variable.

Comparing Individuals

Suppose that the moral performances of Jack and Jill are measured relative to an absolute standard set by an experimenter, along with each of their consistencies. The obtained values for actual performance and consistency are 75% and .9, respectively, for Jack, whereas these values are 70% and .2 for Jill. Who is more moral?

The straightforward answer is that Jack is more moral because his actual performance is at a 75% level compared with 70% for Jill. But appearances can be deceiving. If one uses Equation 6 to correct for attenuation due to inconsistency, then Jack's true moral performance is 76% whereas Jill's is 95%, so Jill's true level of morality is substantially higher than Jack's. This example suggests an interesting philosophical question: What is morality? If one believes that morality should be measured by actual moral performance, then Jack is more moral than Jill. But if one takes the point of view that actual moral performance is only an indirect measure of what is inside the person, then true moral performance would be presumed to be the correct measure of morality and Jill would come off as being superior to Jack. Finally, it is even possible to imagine an argument that behaving consistently with respect to moral issues is what really constitutes morality, in which case Jack's level of .9 clearly outpaces Jill's meager level of .2.

Current Study

The following experiment provides an example of how PPT can be used to analyze individual strategies and consistencies in a moral decision-making task using nonstandard tables. Because PPT is a very abstract theory, an experiment was performed to illustrate more concretely how it can be used.[2] Participants were presented with a series of moral decisions in two blocks separated by a delay task. Although one block is sufficient to obtain a measure of actual performance, using two blocks enabled us to assess consistency across blocks as well. Measures of actual performance and consistency are sufficient to illustrate PPT. We predicted that participants would be less than perfectly consistent in their responses and that their performance could be predicted on the basis of this inconsistency.

Method

Participants

Participants were 30 members of the New Mexico State University community (20 women and 10 men) who were paid $8 per hour for participation. The average age was 20.5 years ($SD = 4.5$).

Materials and Stimuli

The experiment was conducted using a Dell computer with a 2.2 GHz processor and a 20-in. monitor with 1024- × 768-pixel resolution. Participants were seated 20 in. from the monitor, with their position controlled by a chin rest. A standard keyboard was provided for participants' responses. All stimuli were programmed into E-Prime 1.2 (2006).

Procedure

After signing a consent form, participants were seated in a comfortable chair directly facing the experimental display. Instructions were presented onscreen and verbally. Once participants were comfortable with the instructions, they pressed a key to begin the actual experiment.

The experiment consisted of two blocks separated by a delay task. In both blocks, participants were asked to read 50 questions regarding morality choices (ranging in length from 20 to 83 words) and to give *yes* or *no* answers. These questions were presented using standard Times New Roman text with a font size of 14 points. All text was presented in black on a white background. A sample question was

> Suppose that a friend of yours is trying on a dress for a formal. She asks you if you think it looks good on her. You think it does not but also know that if you tell the truth it will hurt her feelings. Are you morally obligated to tell the truth?

[2] We thank our action editor for suggesting that we perform this illustration.

If participants wished to answer *yes*, they were instructed to press the *J* key, and if they wished to answer *no*, they were instructed to press the *F* key. The same exact 50 questions were again presented in Block 2, with the same instructions.

In the delay task, participants were asked to perform a memory task that was designed to be resource demanding to reduce the amount of recall between Blocks 1 and 2. During this task, participants were first presented with a fixation display for 1,000 ms. The fixation display consisted of a black plus sign in the center of a white screen. This screen was then replaced with a number display that consisted of a seven-digit number that appeared in the center of the screen for 5,000 ms. Participants were instructed to memorize this number. Participants were then presented with a count display containing a random number of dots (4–10) and were given 10,000 ms to count the number of dots and give their response by pressing the appropriate key on the number pad. A feedback display was then presented for 1,000 ms, which indicated whether the answer was correct or incorrect. After this, a fresh display appeared asking, "What was the *n*th number?" referring back to the seven-digit number participants had memorized earlier. Participants were given 10,000 ms to respond by pressing the appropriate number on the keypad. Finally, a second feedback display was presented for 1,000 ms, which informed the participant as to whether he or she got the answer correct or incorrect.

All participants were presented with identical scenarios and tasks. On completion of the experiment, participants were debriefed and paid for their participation.

Results

The data were analyzed using a round-robin procedure whereby each participant was compared with each other participant. This resulted in a total of 435 nonredundant pairwise comparisons. A table of observed and true scores was created for each pairwise comparison using the equations presented earlier. Table 2 presents the individual consistencies (r), individual actual successes, individual true scores, A_{DIFF}, B_{DIFF}, C_{DIFF}, D_{DIFF}, and R, with each variable represented as an average of the pairwise comparisons between participant i and the other 29 participants. Composite actual successes and composite true scores are added as well to show composite performance.

The participants were surprisingly consistent. Both the mean and the median consistencies across participants were $r = .80$, ranging from .60 to 1.0. The mean proportion of actual successes was 65% and the mean proportion of true successes was 70%, thereby indicating that if people were perfectly consistent, they would agree with each other 70% of the time rather than 65% of the time, for an average improvement of 5%: Using each participant's average pairwise comparison as a case, $t(29) = 17.5$, $p < .001$. In addition, when the participants were combined into a composite participant (consistency = .88), the average difference between true and observed successes was again equal to 5%, $t(29) = 11.7$, $p < .001$. Thus, even when consistency is quite high, improvement is nevertheless possible by increasing consistency.

Table 2
Actual Successes, True Successes, and Consistency

Participant	r	Individual successes	Individual true scores	A_{DIFF}	B_{DIFF}	C_{DIFF}	D_{DIFF}	R	Composite successes	Composite true scores
S1	.92	.69	.73	0.91	−0.91	−0.91	0.91	.45	.78	.81
S2	.76	.64	.68	1.01	−1.01	−1.01	1.01	.36	.70	.74
S3	.69	.65	.71	1.41	−1.41	−1.41	1.41	.42	.76	.83
S4	.68	.65	.71	1.58	−1.58	−1.58	1.58	.42	.74	.81
S5	.92	.58	.61	0.56	−0.56	−0.56	0.56	.22	.60	.61
S6	.70	.62	.66	0.97	−0.97	−0.97	0.97	.34	.68	.73
S7	.96	.68	.71	0.61	−0.61	−0.61	0.61	.43	.80	.83
S8	.72	.60	.64	1.03	−1.03	−1.03	1.03	.29	.68	.73
S9	.76	.66	.72	1.45	−1.45	−1.45	1.45	.44	.78	.84
S10	.96	.69	.72	0.65	−0.65	−0.65	0.65	.44	.76	.78
S11	.92	.68	.71	0.70	−0.70	−0.70	0.70	.43	.80	.83
S12	.80	.68	.72	1.16	−1.16	−1.16	1.16	.46	.82	.88
S13	.88	.68	.72	0.92	−0.92	−0.92	0.92	.44	.78	.82
S14	.96	.71	.74	0.84	−0.84	−0.84	0.84	.49	.84	.87
S15	.64	.66	.72	1.61	−1.61	−1.61	1.61	.45	.74	.82
S16	.76	.69	.74	1.30	−1.30	−1.30	1.30	.49	.78	.84
S17	.60	.65	.73	1.91	−1.91	−1.91	1.91	.45	.72	.80
S18	.92	.71	.74	0.80	−0.80	−0.80	0.80	.48	.84	.88
S19	.60	.62	.69	1.70	−1.70	−1.70	1.70	.39	.70	.78
S20	.80	.70	.74	1.19	−1.19	−1.19	1.19	.49	.86	.93
S21	.92	.64	.67	0.77	−0.77	−0.77	0.77	.35	.72	.74
S22	.92	.70	.74	0.83	−0.83	−0.83	0.83	.48	.86	.90
S23	.62	.59	.65	1.39	−1.39	−1.39	1.39	.31	.60	.63
S24	.92	.72	.76	0.92	−0.92	−0.92	0.92	.52	.88	.92
S25	.81	.55	.60	1.09	−1.09	−1.09	1.09	.19	.54	.55
S26	.64	.63	.69	1.36	−1.36	−1.36	1.36	.38	.72	.79
S27	.75	.59	.63	0.98	−0.98	−0.98	0.98	.27	.60	.62
S28	.84	.65	.69	0.81	−0.81	−0.81	0.81	.38	.68	.71
S29	.72	.66	.71	1.34	−1.34	−1.34	1.34	.43	.74	.80
S30	.79	.59	.63	1.12	−1.12	−1.12	1.12	.29	.64	.67
M	.797	.653	.697	1.098	1.098	1.098	1.098	.400	.738	.784

Although our purpose was to illustrate how the theory can be used, some interesting predictions can be tested even with the present simple experiment. For example, because the calculation of true successes takes consistency into account, consistency should not be highly correlated with true successes (although the correlation might not equal zero if, e.g., people with better strategies also tend to be more consistent), whereas consistency should be more correlated with actual successes and especially correlated with the difference between actual and true successes. In fact, the findings supported these predictions. Consistency was not significantly correlated with true successes ($r = .22$, $p > .2$), it was moderately correlated with actual successes ($r = .46$, $p = .01$), and it was highly correlated with the difference between true and actual successes ($r = .80$, $p < .001$).

An alternative procedure for analyzing the importance of consistency is to assess the 30 pairwise comparisons of each participant against the composite. In this case, the consistency of the composite is a constant (it equals .88) and the question is how much the consistencies of the participants correlate with actual successes, true successes, and the difference between actual and true successes. As in the previous analyses, consistency is not particularly correlated with true performance ($r = .21$, $p > .2$), it is moderately correlated with actual performance ($r = .47$, $p = .01$), and it is extremely correlated with the difference between actual and true performance ($r = .86$, $p < .001$).

One potential benefit of PPT is that it can be used to answer questions that come from outside the theory. Consider, for instance, the issue of why composite performances are often better than the average individual performances on a variety of tasks. The usual explanation is that when a composite is formed, the errors compensate for each other, thereby resulting in better performance. An example of this would be when people guess the number of beans in a jar. Some people overestimate whereas other people underestimate, and so the composite (average) score tends to be more accurate than most of the individual guesses.

An additional possibility, however, in the context of moral performance, is that the formation of a composite person actually results in the capture of aspects of each person's strategy. This latter possibility suggests that the composite person should have a higher level of performance than the average individual person, even after correcting for attenuation due to inconsistency. To test these possibilities, we performed a 2 (participant type: composite person vs. real people) $\times$ 2 (performance type: actual vs. true) analysis of variance (ANOVA) on the correlations. Not surprisingly, both main effects were significant, indicating that true performance was better than actual performance (Ms = 74% and 69.5%, respectively), $p < .001$, and that the composite person outperformed the average individual person (Ms = 76% and 67.5%, respectively), $p < .001$.

More important, although the composite person was more consistent than the average individual person (M consistencies = .88 and .80, respectively), there was no interaction, $F(1, 29) < 1$. The original advantage for the composite person over the average individual person in actual performance (Ms = 74% and 65%, respectively) was retained even when true scores were evaluated (Ms = 78% and 70%, respectively). If the advantage for the composite person in actual performance had been due solely to greater consistency, then this advantage should have been eliminated when true scores were computed.

Discussion

Until this point, our discussion has been restricted to explicating PPT, showing what it can do, and providing a simple experimental demonstration. But other issues remain to be discussed. First, an important aspect of PPT is that it is easy to integrate with other theories; an example of this is in the context of signal detection theory (Green & Swets, 1966; Trafimow & Rice, 2008), because signal detection theory is also based on a 2 $\times$ 2 performance matrix. Second, there are various possible interpretations of PPT that we do not favor and these are discussed in an attempt to prevent confusion. Third, although our focus thus far has been abstract and theoretical, PPT can be applied to many areas other than morality, and some of these are mentioned. Fourth, we mention some disadvantages and advantages of PPT, including some possible avenues for addressing the disadvantages in future theorizing.

What Is Not Being Proposed

Our concentration on strategies and the consistencies with which they are used might be interpreted to imply positions that were unintended. For example, our focus on moral strategies should not be taken as an argument against information processing theories, such as social information processing theory (e.g., Crick & Dodge, 1994; Dodge & Rabiner, 2004) or moral attribution theory (e.g., Reeder & Brewer, 1979; Trafimow, Bromgard, Finlay, & Ketelaar, 2005; Trafimow & Trafimow, 1999). That there are various steps in the process by which people make moral decisions, such as encoding, interpretation, exploring options, making goals, and so on, has been well documented, although to what extent the steps happen sequentially, in parallel, or in some other way is not yet completely clear.

For our purposes, what counts is that these steps eventually culminate in a response decision (e.g., Step 5 in Crick & Dodge, 1994) and possibly a behavior (see Wyer & Srull, 1989, for a general view of social cognition processes). To the extent that the steps outlined in social information processing theory or other social cognition theories are not random, they constitute a strategy to which PPT is fully applicable. Thus, consistent with the call that Arsenio and Lemerise (2004) made for social information processing and moral domain models to inform each other, PPT is capable of incorporating both approaches under a large, although admittedly abstract, umbrella.

In addition, PPT should not be taken as making a commitment to the trait–situation debate (Allport, 1955; Block, 1977; Epstein, 1979, 1980, 1983, 1984; Kenrick & Funder, 1988; Mischel, 1968; Mischel & Peake, 1982; see Pervin, 2002, for a recent review). It might be tempting to liken strategies to individual differences and randomness to situations and to interpret PPT as favoring the former over the latter. However, matters are more complicated than that. For one thing, it is far from clear that strategies should be considered to be an individual differences variable. Clearly, people use different strategies in different situations and so, to put it in traditional terms, there is an interaction between individual differences and situations. Further, it is not a one-time ANOVA type of interaction because people's internal characteristics influence the situations they experience, which, in turn, influence how they are, and so on in a never-ending spiral. From the perspective of PPT, the totality of all of these factors comprises a compound strategy that cannot be separated easily into individual differences, situations, and/or a one-time ANOVA type of interaction.

In contrast to implying that situations do not matter, PPT can actually be used as a tool to decompose compound strategies into the specific strategies people use in different situations. To see this, consider Experimenter A, who performs a PPT experiment across a variety of situations. The result will give a good account of the true efficacy of the compound strategies used by the participants. But suppose that Experimenter B is interested in decomposing participants' compound strategies into the strategies that they use in particular situations. This can be done by presenting participants with scenarios that are blocked by situation type and running PPT analyses within each situation type. Thus, PPT not only is capable of incorporating theories about situations and about Person $\times$ Situation interactions but can also be used to guide research that pertains to them.

Applying PPT

One advantage of the abstract nature of PPT is that the mathematics is sufficiently flexible to handle performance as evaluated by either absolute or relative standards. Clearly, the application of PPT is more straightforward when absolute standards are relevant, such as in the types of visual cognition tasks favored by many psychologists. For example, when participants claim to recognize that a stimulus has been presented before, there is a clear, objective answer. Likewise, some engineering psychologists (and applied cognitive psychologists) tend to be interested in tasks with security implications such as determining whether luggage contains weapons, whether an enemy target is in a designated area, and so on—all tasks with objectively correct answers. Even going outside of psychology, many tasks can be assessed objectively: How well do stockbrokers predict stock prices, how well do gamblers play poker hands, how accurately do doctors diagnose the flu? In all of these cases, it should be easy to obtain measures of actual performance and consistency. Consequently, PPT can be applied straightforwardly and with full force.

Also interesting are the less straightforward cases where there is no absolute standard. Consider, for example, the plight of clinical psychologists who specialize in marriage counseling and are faced with couples who disagree with each other on countless occasions. Suppose that the Smiths agree 60% of the time and disagree 40% of the time. It could be that both people are very consistent in their choices, in which case disagreement could be reduced only by having someone change his or her worldview, by having someone change his or her way of making decisions, or by trying some other type of intervention that is difficult for the therapist to accomplish. Or it could be that one or both of the people are very inconsistent, in which case PPT suggests that an effective intervention might involve increasing the consistency of Mr. or Mrs. Smith, possibly via the use of self-affirmation, rehearsal, or other treatments.

Ironically, when therapists present their clients with alternative strategies, the effect might be to decrease consistency and thereby increase disagreements! Our point here is not that therapists should never present alternative strategies. Rather, we believe that it would be useful to perform a PPT analysis prior to intervention to determine whether the therapist would be better off focusing on consistency or on alternative strategies. For those couples whose actual and true relative performances are substantially different, therapists should consider a consistency-based treatment, whereas a strategy-based treatment would be more likely to be effective when actual and true relative performances are similar.

Of course, marital therapy is only one area where relative performance is what counts. Others might include agreement among administrative teams, among PhD committees, and so on. PPT might even be applied to team sports. For example, there might be many ways for National Football League receivers to run good pass patterns; what matters is that the way a pattern is run is in agreement with how the quarterback expects that pattern to be run. Similar comments could be made about how soccer or hockey players link up, how keystone combinations lead to double plays, and other sports contexts.

Increasing the Number of Choices

One of the main limitations of the present approach is that it is based on the idea of 2×2 tables. But what if one is interested in cases involving three or more choices? At present, PPT has not been developed to the point where it can handle such cases in a convenient way. But we believe that PPT can be developed in that direction. Although this development requires its own article, an outline of the likely trend is presented in this section, along with a discussion of some of the philosophical issues that such an expansion of PPT would bring about.

As always, the first step in dealing with multiple-choice scenarios would be to convert the observed table of outcomes into some kind of coefficient of association. Here we run into our first philosophical problem. Specifically, are the choices ordered or not? In the case of relative task performance, are all disagreements between people equal, or do some kinds of disagreements indicate more distance between strategies than do other kinds of disagreements? Or, in the case of absolute task performance, are all wrong choices equally wrong or are some less wrong than others? If we assume that there is no ordering, then it makes sense to use the Cramér coefficient or some derivative of it (see Siegel & Castellan, 1988, pp. 225–232). If the choices can be ordered, then it makes more sense to use some kind of rank-order coefficient. It may be necessary to collect data on people's perceptions of the ordering of task performance options in various contexts before it will be possible to choose an appropriate vehicle for converting multioption tables into coefficients of association.

Even after the foregoing problem is solved, a second problem of adjusting the correlation for attenuation due to inconsistency remains. How does one measure consistency when there are multiple options that may or may not be ordered? Again, the issue of ordering must be addressed before this problem can be solved. Finally, even after these two problems are addressed, there remains the problem of converting the true coefficient of association back into either a standard or a nonstandard table. In the case of multiple choices, it is not necessarily clear what a standard table would look like. In principle, it should at least be possible to create nonstandard tables, but there is an important mathematical impediment. Specifically, as the number of choices increases, so does the number of unknown variables, thereby necessitating more equations to find their values. None of these problems seem unsolvable, but some prior conceptual work and possibly some prior empirical work will be necessary.

Disadvantages and Advantages

Like any theory, PPT has disadvantages and advantages. It is particularly important to be candid about the disadvantages, and so these will be addressed first.

Disadvantages. It is impossible to have a theory without assumptions and, by their very nature, assumptions tend to be unproven. This is true of PPT as well. The first assumption, that the decisions of concern involve only two choices, clearly limits the range of applicability of the theory. Although in many situations there really are only two choices (e.g., Kohlberg's moral scenarios)—or only two choices that are routinely accessible to the person—other situations have several accessible choices. As the foregoing section indicates, we believe that PPT will be developed to the point where it will address multichoice cases and that the present theory will actually be a special case of the more general theory. But until that time, we are forced to acknowledge the theory's limited range of applicability.

The second assumption is that the assumptions of classical true score theory are true, or at least good enough. Although, as Cohen and Swerdlik (1999) pointed out, classical true score theory has dominated the field for many decades, continues to do so, and is still accepted routinely, it is possible that its assumptions may nevertheless be incorrect. At the present time, there is no way to determine this with perfect confidence. The good news, however, is that PPT makes very precise predictions (we say more about this later), and so it should be extremely easy to falsify if it is wrong. If PPT turns out to be wrong, then, by implication, the assumptions of classical true score theory, on which it is based, would also be wrong. Such a discovery would overturn large areas of psychology where research is based on the assumptions of classical true score theory. Consequently, at worst, research that disconfirms PPT would also cause psychologists to reexamine many of their basic assumptions, which in itself would be a valuable contribution. At best, PPT's predictions would be confirmed, which would both increase understanding of morality and also gain the chance for useful applications that would increase the goodness of the world.

There are also empirical difficulties. Possibly the most important difficulty concerns the measurement of actual performance and consistency. The easy way to do this would be to present people with scenarios and measure actual performance and consistency via their responses to these scenarios. But researchers might disagree about whether people's responses to scenarios would parallel what they would do in real life—the classic issue of generality of findings. For those researchers who are less concerned with generality, scenarios might be fine. But those researchers who are more concerned with it might have to resort to more complicated methods such as diaries, reports by friends or significant others, or even direct observation. It is hoped that the results of the different methods would converge on similar conclusions, but they might not, and it is important to be open to this possibility.

Advantages. The most obvious advantage of PPT is that it is parsimonious—the number of assumptions and variables is kept to a minimum. This reduces conceptual confusion and makes PPT easy to test and easy to use. The extreme parsimony of PPT compares favorably with other mathematical theories that contain large numbers of free parameters. Clearly, when there are many free parameters, it is easy to make seemingly small adjustments here or there that allow the theory to predict (or postdict) almost any set of findings (see Roberts & Pashler, 2000, for an excellent discussion of this issue). The present theory, with its paucity of variables, is less vulnerable to this criticism.

A limitation of nonmathematical theories is that their predictions tend to be very imprecise. Thus, any obtained data can be argued to either support or disconfirm them, with much of the argument devolving to what additional assumptions people wish to make. In contrast, the present theory makes extremely precise predictions that are much less susceptible to add hoc arguments. This extreme precision makes the theory easy to test and use. The predicted values will be confirmed or disconfirmed by data, and the virtues or faults of the theory will be apparent to everyone.

There is yet an additional, though subtle, advantage. Many decades ago, Gordon Allport (1937) pointed out that much research in psychology is subject to the fallacy of the group mind. In essence, Allport argued that psychologists calculate group means or group correlations that may have nothing at all to do with any particular person within the group. Nevertheless, researchers draw conclusions about people, often without considering this issue. The present theory, because it makes precise predictions about each individual, is much less likely than other theories to promote this fallacy.

Conclusion

Many psychological variables contribute to task performance, such as accessible concepts, affects, and decision strategies. Our claim is not that PPT is capable of replacing all of these other variables. However, we do claim that if those other variables influence task performance, they do it by influencing the goodness of people's strategies, consistency, or both. Thus, PPT should not be considered to be a complete theory of task performance but rather a proximal theory that leaves plenty of room for the distal variables that have been the concern of philosophers and psychologists over the millennia. Future research efforts might profitably be devoted to testing how traditional psychological variables influence consistency and strategy, which, in turn, influence task performance.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Monterey, CA: Brooks/Cole.

Allport, G. (1937). *Personality: A psychological interpretation.* New York: Holt.

Allport, G. (1955). *Becoming: Basic considerations for a psychology of personality.* New Haven, CT: Yale University Press.

Aristotle. (2000). *The Nicomachean ethics* (R. Crisp, Ed. & Trans.). Cambridge, UK: Cambridge University Press. (Original work c. 330 BCE)

Arsenio, W. F., & Lemerise, E. A. (2004). Aggression and moral development: Integrating social information processing and moral domain models. *Child Development, 75,* 987–1002.

Block, J. (1977). Advancing the psychology of personality: Paradigmatic shift or improving the quality of research? In D. Magnusson & N. Endler (Eds.), *Personality at the crossroads* (pp. 37–64). Hillsdale, NJ: Erlbaum.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurement* (4th ed.). Mountain View, CA: Mayfield.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74–101.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth/Thomson Learning.

Cronbach, L. J., Gleser, G. C., & Rajaratnam, N. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Mathematical and Statistical Psychology, 16,* 137–173.

Dodge, K. A., & Rabiner, D. L. (2004). Returning to roots: On social information processing and moral development. *Child Development, 75,* 1003–1008.

E-Prime (Version 1.2). (2006). [Computer software]. Pittsburgh, PA: Psychology Software Tools.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37,* 1097–1126.

Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist, 35,* 790–806.

Epstein, S. (1983). A research paradigm for the study of personality and emotions. In M. M. Page (Ed.), *Personality: Current theory and research* (pp. 91–154). Lincoln: University of Nebraska Press.

Epstein, S. (1984). The stability of behavior across time and situations. In R. A. Zucker, J. Aronoff, & A. I. Rabin (Eds.), *Personality and the prediction of behavior* (pp. 209–268). Orlando, FL: Academic Press.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development.* Cambridge, MA: Harvard University Press.

Gilligan, C., & Attanucci, J. (1988). Two moral orientations: Gender differences and similarities. *Merrill-Palmer Quarterly, 34,* 223–237.

Green, D. M., & Swets J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814–834.

Hardin, G. (1982). Discriminating altruisms. *Zygon, 17,* 163–186.

Hobbes, T. (1957). *Leviathan* (M. Oakeshott, Ed.). Oxford, England: Oxford University Press. (Original work published 1651)

Hume, D. (1978). *A treatise of human nature* (2nd ed.; L. A. Selby-Bigge & P. H. Nidditch, Eds.). Oxford, England: Clarendon Press. (Original work published 1740)

Johnston, V. S. (1999). *Why we feel: The science of human emotions.* Reading, MA: Perseus Books.

Kant, I. (1991). *The metaphysics of morals* (M. Gregor, Trans.). Cambridge, England: Cambridge University Press. (Original work published 1797)

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person–situation debate. *American Psychologist, 43,* 23–34.

Kohlberg, L. (1981). *The philosophy of moral development: Essays on moral development* (Vol. I). San Francisco: Harper & Row.

Kohlberg, L. (1984). *The psychology of moral development: Essays in moral development* (Vol. II). San Francisco: Harper & Row.

Korsgaard, C. M. (1985). Kant's formula of universal law. *Pacific Philosophical Quarterly, 66,* 24–47.

Martin, G. (1989). The utility of speech input in user–computer interfaces. *International Journal of Man–Machine Systems Study, 18,* 355–376.

Mill, J. S. (1979). *Utilitarianism* (G. Sher, Ed.). Indianapolis, IN: Hackett. (Original work published 1861)

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Mischel, W., & Peake, P. K. (1982). Beyond déjà vu in the search for cross-situational consistency. *Psychological Review, 89,* 730–755.

Pervin, L. A. (2002). *Current controversies and issues in personality* (3rd ed.). New York: Wiley.

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86,* 61–79.

Roberts, S., & Pashler, S. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis.* New York: McGraw-Hill.

Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology, 9,* 395–396.

Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166–169.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw-Hill.

Smith, A. (1976). *The theory of moral sentiments* (D. D. Raphael & A. L. Macfie, Eds.). Oxford, England: Clarendon Press. (Original work published 1759)

Thomson, J. J. (1989). The no reason thesis. *Social Philosophy and Policy, 7,* 1–21.

Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review, 110,* 526–535.

Trafimow, D., Bromgard, I. K., Finlay, K. A., & Ketelaar, T. (2005). The role of affect in determining the attributional weight of immoral behaviors. *Personality and Social Psychology Bulletin, 31,* 935–948.

Trafimow, D., & Rice, S. (2008). *Integrating potential performance theory (PPT) and signal detection theory (SDT).* Manuscript submitted for publication.

Trafimow, D., & Sheeran, P. (2004). A theory about the translation of cognition into affect and behavior. In G. Maio & G. Haddock (Eds.), *Contemporary perspectives in the psychology of attitudes: The Cardiff Symposium* (pp. 57–75). London: Psychology Press.

Trafimow, D., & Trafimow, S. (1999). Mapping imperfect and perfect duties onto hierarchically and partially restrictive trait dimensions. *Personality and Social Psychology Bulletin, 25,* 686–695.

Wyer, R. S., & Srull, T. K. (1986). Human cognition in its social context. *Psychological Review, 93,* 322–359.

Wyer, R. S., & Srull, T. K. (1989). *Memory and cognition.* Hillsdale, NJ: Erlbaum.

Appendix A

Proof for Equation 5

As presented earlier, Equation 5 appears below.

$$S = \frac{(2s - 1 + \sqrt{r_{xx'}r_{yy'}})}{2\sqrt{r_{xx'}r_{yy'}}}. \quad (5)$$

Equations 2 and 3 are explained in the main text and are represented as assumptions for this proof as Equations A1 and A2, respectively.

$$R = \frac{r}{\sqrt{r_{xx'}r_{yy'}}}. \quad (A1)$$

$$S = \frac{(R + 1)}{2}$$

in terms of true scores, or, in terms of actual scores,

$$s = \frac{(r + 1)}{2}. \quad (A2)$$

A simple rearrangement of the terms in the actual score version of Equation A2 results in Equation A3.

$$r = 2s - 1. \quad (A3)$$

Equation A3 can be substituted into Equation A1, thereby resulting in Equation A4.

$$R = \frac{(2s - 1)}{\sqrt{r_{xx'}r_{yy'}}}. \quad (A4)$$

Finally, substituting Equation A4 into Equation A2 (the true score version) results in Equation 5.

Appendix B

Proofs for Equations 9–12

As presented earlier, Equations 9–12 appear below.

$$A = \frac{R\sqrt{R_1R_2C_1C_2} + C_1R_1}{(R_1 + R_2)}. \quad (9)$$

$$B = \frac{R_1(R_1 + R_2) - (R\sqrt{R_1R_2C_1C_2} + C_1R_1)}{(R_1 + R_2)}. \quad (10)$$

$$C = \frac{C_1R_2 - R\sqrt{R_1R_2C_1C_2}}{(R_1 + R_2)}. \quad (11)$$

$$D = \frac{C_2(R_1 + R_2) - [R_1(R_1 + R_2) - (R\sqrt{R_1R_2C_1C_2} + C_1R_1)]}{(R_1 + R_2)}. \quad (12)$$

We start this proof with Equation 1 that was explained in the main text and is repeated below as Equation B1.

$$r_\phi = \frac{|ad - bc|}{\sqrt{r_1r_2c_1c_2}} = \frac{|ad - bc|}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}. \quad (B1)$$

Also, in the main text, we pointed out that once the observed correlation was obtained, it could be converted into a true correlation via Equation B2 below.

$$R = \frac{r}{\sqrt{r_{xx'}r_{yy'}}}. \quad (B2)$$

Given that the observed correlation obtained from an observed table has been converted into a true correlation, the task now is to convert the true correlation back into a true table. Thus, we assume first that Equation B1 can be expressed in true score terms as well as in observed score terms. And second, in contrast to when we were concerned with standard tables such as in Appendix A, we also wish to preserve the margin frequencies from the observed table. Thus, in contrast to Appendix A, we assume Equations B3–B7.

$$R = \frac{|AD - BC|}{\sqrt{R_1R_2C_1C_2}} = \frac{|AD - BC|}{\sqrt{(A + B)(C + D)(A + C)(B + D)}}, \quad (B3)$$

$$R_1 = r_1, \quad (B4)$$

$$R_2 = r_2, \quad (B5)$$

$$C_1 = c_1, \text{ and} \quad (B6)$$

$$C_2 = c_2. \quad (B7)$$

Also, by definition, we assume that the sum of the cell frequencies in each row or column equals the row or column frequency,

(*Appendixes continue*)

respectively. Thus, Equations B8–B11 can be derived directly from the definitions.

$$R_1 = A + B, \tag{B8}$$

$$R_2 = C + D, \tag{B9}$$

$$C_1 = A + C, \text{ and} \tag{B10}$$

$$C_2 = B + D. \tag{B11}$$

Rearranging Equation B3 leads to Equation B12.

$$R\sqrt{R_1R_2C_1C_2} = AD - BC. \tag{B12}$$

Rearranging Equations B8–B11 leads to Equations B13–B16.

$$B = R_1 - A \text{ (from B8)}, \tag{B13}$$

$$D = R_2 - C \text{ (from B9)}, \tag{B14}$$

$$C = C_1 - A \text{ (from B10), and} \tag{B15}$$

$$D = C_2 - B \text{ (from B11)}. \tag{B16}$$

Equations B13 and B14 can substitute for B and D, respectively, which eliminates two unknowns in Equation B17.

$$R\sqrt{R_1R_2C_1C_2} = [A(R_2 - C) - C(R_1 - A)]. \tag{B17}$$

Equation B18 has two unknowns left (A and C) and one of these (C) can be eliminated by substitution using Equation B15, thereby rendering Equation B18.

$$R\sqrt{R_1R_2C_1C_2} = [A(R_2 - (C_1 - A))] - [(C_1 - A)(R_1 - A)]. \tag{B18}$$

Expanding Equation B18 leads to Equation B19.

$$R\sqrt{R_1R_2C_1C_2} = R_2A - C_1A + A^2 - C_1R_1 + C_1A + R_1A - A^2. \tag{B19}$$

Some subtraction on the right side of Equation B19 renders Equation B20.

$$R\sqrt{R_1R_2C_1C_2} = R_2A - C_1R_1 + R_1A. \tag{B20}$$

It is now necessary to isolate A by keeping all of the terms with it on one side of the equation as in Equation B21.

$$R\sqrt{R_1R_2C_1C_2} + C_1R_1 = R_2A + R_1A. \tag{B21}$$

Factoring to further isolate A renders Equation B22.

$$R\sqrt{R_1R_2C_1C_2} + C_1R_1 = A(R_2 + R_1). \tag{B22}$$

Finally, dividing both sides by $R_2 + R_1$ renders Equation B23 below, which is equivalent to the first goal: Equation 9.

$$\frac{R\sqrt{R_1R_2C_1C_2} + C_1R_1}{(R_1 + R_2)} = A. \tag{B23}$$

With A solved, obtaining B, C, and D is easy. Substitution of Equation B23 into Equation B13 solves for B, substitution of Equation B23 into Equation 15 solves for C. And once B is solved, substitution of that solution into Equation B16 solves for D. Thus, the problem of obtaining Equations 9–12 is solved.

Received February 22, 2007
Revision received January 21, 2008
Accepted January 22, 2008 ■

Correction to Sedlmeier and Kılınç (2004)

The correction notice to "The Hazards of Underspecified Models: The Case of Symmetry in Everyday Predictions," by Peter Sedlmeier and Berna Kılınç (*Psychological Review*, 2004, Vol. 111, No. 3, pp. 770–780), which printed in Vol. 115, No. 1., p. 198, contained an incorrect DOI. The correct DOI is as follows: 10.1037/0033-295X.115.1.198

DOI: 10.1037/0033-295X.115.2.462