THE SOCIAL CONSTRUCTION OF CARDIOVASCULAR REACTIVITY

Nicholas Christenfeld, Ph.D., Laura M. Glynn, M.A., and James A. Kulik, Ph.D.
University of California, San Diego

William Gerin, Ph.D.
Cornell University Medical College

ABSTRACT

One can view the "reactivity hypothesis" as having two basic forms: the individual difference or personality approach, which suggests that people who show exaggerated cardiovascular responses to stress are at increased risk of developing cardiovascular disease, and the situational or social psychological approach, which suggests that circumstances which give rise to unusually large responses are those that put people at risk of disease. Both versions rely on the generality of cardiovascular responses across situations. Evidence is presented from two studies which indicate that such generality may, however, be hard to come by. In the first study, examining the personality approach, we show that a simple change in setting dramatically attenuates the consistency of reactivity. In the second study, from the social psychological perspective, we show that subtle alterations in the situation have profound effects on group mean responses. In both cases, reactivity proved extremely sensitive to the context, suggesting that testing in arbitrary and artificial settings cannot be expected to generalize well to the real world. Instead, we argue, careful attention to psychological naturalism is essential, with the testing carefully matched to specific real-world phenomena of interest.

(**Ann Behav Med** 1998, *20*(4):317–325)

INTRODUCTION

There is now a large body of work devoted to exploring cardiovascular responses to stress. Much of this is based on the "reactivity hypothesis." We suggest it is possible to divide the reactivity hypothesis into two forms: a "personality" version and a "social psychology" version. The personality version argues that one can identify people who are at elevated risk of future hypertension or heart disease on the basis of exaggerated cardiovascular responses to stress. That is, reactivity is a risk factor for cardiovascular disease (CVD) (1–5). The personality approach has attracted the bulk of the attention in the field and traditionally has been divided into two forms—the "strong" form and the "weak" form (1,2). The strong form is based on the notion that exaggerated blood pressure and heart rate responses to stress damage the cardiovascular system, and so people who tend to show such responses are at increased risk of developing CVD. The weak form suggests exaggerated responses may not be causally implicated in the development of hypertension or heart disease, but instead may serve as a marker for future disease. In either event, reactivity is an individual difference measure—used to classify people as those at higher or lower risk of future disease.

The social psychological approach to the reactivity hypothesis suggests that tasks or circumstances which produce large cardiovascular responses may be damaging to the system. Similarly, manipulations or interventions that reduce reactivity may be beneficial. This view of reactivity is necessarily causal. People are at risk of future heart disease because of wide blood pressure excursions, and limiting those will reduce that risk. This approach does not depend on individual differences. The point of the social psychological approach is not to identify people at risk, but instead to identify situations that put people at risk or to identify interventions that reduce that risk.

There is evidence for both the personality and social psychological views, though neither is beyond controversy. Cynamolgus macaques, for example, who are classified as high reactors show more atherosclerosis after repeated stressor episodes than do low-reactor macaques (6,7). These data support the personality view of the hypothesis. Furthermore, some macaques who are subjected to stressors in the form of unstable social environments also show signs of cardiovascular damage (8,9). This evidence for the causal role of reactivity supports the social psychological approach and also the strong personality model.

In humans, there is also evidence for the personality view: some studies, though not all (1), indicate that people who show larger blood pressure responses to stress are more likely to develop hypertension (2–5,10). Along similar lines, blood pressure reactivity to mental stress has been shown to predict increase in atherosclerosis over the course of two years (11). In addition, other individual differences, particularly anger and hostility, have also been linked to both reactivity (12,13) and to the development of coronary heart disease (CHD) (14,15). This evidence for the personality view does not address the causal role of reactivity. It is possible that individuals who show the greater cardiovascular responses to stress are those who, due to other factors, are also predisposed to develop CVD. In such a case, the observed exaggerated blood pressure and heart rate responses may be epiphenomenal, but still may be useful markers for CVD. It is not the purpose of the present article, however, to review the state of the evidence for the link between reactivity and disease. It is, instead, the intent of this article to explore the generalizability of cardiovascular reactivity and the limitations that low generalizability imposes on both the personality and social psychological approaches to reactivity.

Both the personality and the social psychological views of reactivity depend, at least indirectly, on there being some generalizability of the cardiovascular responses. The two approaches, however, depend on different sorts of generalizability. The personality approach requires that the measurement of an individual's reactivity to stress measured in one context predicts that person's

Reprint Address: N. Christenfeld, Ph.D., Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109.

response in other contexts. The presence of person-by-situation interactions is troublesome, because there cannot be "reactors" who go about overresponding to stress if there is no cross-situational consistency in the extent to which people respond to stress. In other words, the concept of reactors implies that people who are identified in one context as high reactors ought to be also the high reactors when they are measured in another context. For example, people might be rank-ordered in terms of how much their blood pressure is elevated above baseline during a laboratory cold pressor task. If these blood pressures are to be predictive of later disease, it ought to be the case that at least roughly the same rank order of reactivity is preserved in the reactions to everyday stressors. The people who show large responses to the cold pressor should be the same people who show large responses when they are stuck in traffic jams, waiting for the dentist to repair a filling, or dealing with a recalcitrant clerk. If reactivity does not generalize from one situation to another, then reactivity in any one setting is unlikely to be predictive of later disease.

The person-by-situation interaction is troubling in slightly different ways for the weak and strong versions of the personality approach. For the weak form, if reactivity scores are different in each setting, it may not be possible to know which setting will produce scores predictive of disease. Will it be response to a cold pressor that predicts later disease, or response to anger provocation, or to mirror tracing, mental arithmetic, or any of a number of other stressors? The strong version of the personality approach to reactivity likewise depends critically on there being cross-situational consistency in an individual's reactivity, because it is the accumulation of such episodes that is hypothesized to produce the disease end state. If response to the cold pressor does not predict responses to other stressors, then high cold pressor reactors will not, at the end of the average day, have shown any more reactivity than low cold pressor responders and so will not have done any more damage to their cardiovascular system. It is therefore the case that the prevalence of the person-by-situation interaction in the assessment of reactivity is a limiting factor in the individual difference approach.

The social psychological approach to reactivity depends on a different sort of generalizability. Because this approach does not rely on individual differences, it is not critical that the rank order of people be preserved from one situation to another. However, it does require that the rank order of situations be preserved. That is, if some situations, such as anger provocation, are damaging because they produce high reactivity, then it needs to be the case that anger provocation generally has this effect. Another way to regard this generalizability issue is that manipulations must generalize from one situation to another. For example, if, in a given setting, one finds that expressing anger reduces blood pressure compared to keeping it contained, then it ought to be the case that the anger expression also will be associated with lower reactivity in another setting. Likewise, if one is to say that social support exerts a beneficial effect because it attenuates reactivity under stress, then it ought to be the case that social support generally has this effect and that the finding is not particular to that one operationalization. The limiting factor is the manipulation-by-situation interaction. That is, receiving advice from a wife may lower reactivity in the supermarket, but raise it when driving on the freeway. This would be an interaction of spousal support by situation. Similarly, having a job with high demand might increase reactivity in group settings, but lower it in solitary tasks; or raise it when there is time pressure, but lower it when there is none. This would be a demand-by-situation interaction. These sorts of interactions prevent one from concluding anything general about what situations (advice from a wife, having a demanding job, etc.) produce reactivity—and thus what situations are potentially harmful. Viewed another way, it prevents one from determining what interventions are beneficial (informational support from a wife, reducing job demands, etc.).

The generalizability of reactivity, in both the personality and social psychological sense, relies on a lack of interactions. The limiting condition for the personality view is the person-by-situation interaction. It is weakened to the extent that the measure of the person is only a measure of the person in that one context and does not predict to other contexts. The limitation of the social psychological approach is the manipulation-by-situation interaction. This is weakened to the extent that the effects of some manipulations are particular to one setting and it cannot be expected to have the same effect in other settings. The existence of these two sorts of interactions is at least logically independent. That is, there can be no consistency in individual differences at the same time that there is considerable stability in the effects of interventions. So, for example, how much reactivity people show when doing a task without an evaluative audience may not predict how much reactivity they show in the presence of such an audience. Nonetheless, it could be that an evaluative audience reliably increases arousal. To take a different sort of example, people may quite reliably wear darker clothes at funerals than at the beach without it being the case that those dressed most darkly at the funeral were also the ones wearing the darkest clothes at the beach. In these cases, there would be person-by-situation interactions, but no manipulation-by-situation interactions. Similarly, it is possible for there to be cross-situational consistency of individual differences without there being stability in the effects of an intervention. For example, providing performance feedback might lower reactivity in some settings (say, when the task is easy), but raise reactivity in others (when the task is difficult). However, the people who show the most reactivity to the task might be the same across the various settings. As another example, drinking whiskey may sometimes make people more violent and sometimes make them less violent. However, it could be that the people who are most violent when drunk are the same ones who are most violent when sober. In these cases, there would be manipulation-by-situation interactions, but no person-by-situation interactions.

There has been a great deal written about the prevalence and implications of the person-by-situation interaction. Hartshorne and May (16), for example, collected data on the cross-situational consistency of honesty in children and found little evidence of generalizability from one setting to another, even when the change was as subtle as the difference between cheating off the answer key on a test and cheating while grading their own tests. Mischel (17), reviewing an enormous amount of evidence from clinical and experimental psychology, concluded that although people are often highly consistent from one time to another, predicting behavior from one situation to another is extremely difficult. He wrote, "In spite of methodological reservations, however, it is evident that the behaviors which are often construed as stable personality trait indicators actually are highly specific and depend on the details of the evoking situations and the response mode employed to measure them" (17, p. 37). Based on his review of the evidence, he suggested that correlations between behaviors across settings rarely exceed a ceiling of $r = 0.3$. Nisbett and Ross (18), reviewing the literature more recently, reach similar conclusions.

The view that personality characteristics are not consistent across situations has been controversial. It has been suggested, for

example, that people are not consistent at the level of individual behaviors, but that larger aggregations of behavior are consistent (19). It has also been suggested that not all people are consistent on all personality dimensions, but that a more idiographic approach will find that some people are consistent on some traits (20). Still others have argued that cross-situational consistency will emerge not at the level of behavior, but rather at the level of the underlying meaning of the behavior (21). Regardless of whether or not such approaches turn out to be successful, it seems clear that the cross-situational consistency of a trait or behavior cannot be taken for granted and may be quite modest.

The findings from personality psychology have important implications for cardiovascular reactivity. An individual's level of reactivity could be viewed as a pure physiological parameter. People's height, weight, and left ventricular hypertrophy can be measured in any of a number of ways and are stable characteristics of the individual. The measurement of such traits should not depend much on the social context; person-by-situation interactions should be largely irrelevant. However, if reactivity is seen as a physiological expression of a personality dimension—a psychophysiological behavior—then such generality cannot be assumed. A person's reactivity to a stressor will not simply be a function of that person's innate level of reactivity, but will instead depend on their construal of the situation. Reactivity to the cold pressor may be an index of the way an individual confronts a challenge, if the task is presented as something to be overcome. On the other hand, it may tap an individual's response to pain if it is presented as something simply to be endured. These subtle alterations may determine to what other contexts the response generalizes. Furthermore, to the extent that people vary in the way they construe the task, the extent to which their responses can predict to any other single situation will be limited.

The limited cross-situational consistency of personality dimensions, viewed from another perspective, may be interpreted as evidence that people are highly sensitive to even subtle changes in contexts. This notion is the underpinning for such theories as Festinger's theory of social comparison processes (22), which suggests that people do not simply express their personality across situations, but instead modify their beliefs and opinions to conform to the immediate social context.

People's sensitivity to situations will, as Mischel and others have pointed out, limit the cross-situational consistency of personality traits. However, it will also have another effect, which is to limit the cross-situational consistency of interventions or manipulations. An alteration that produces one effect in one context may be interpreted quite differently and produce quite a different effect in another context. To take an extreme example, it may be reassuring to be given control over one's pain medication supply after surgery, but not reassuring to be given control over one's jumbo jet coming in on final approach. Similarly, adding an audience may in some contexts make people perform better and in some perform worse (23). When these interactions are understood and stable, they can be incorporated into the theory. However, there are many dimensions over which effects can interact, and one cannot presume stability of an effect from one context to another.

We will now turn the discussion to two sets of empirical findings that illustrate the context sensitivity of cardiovascular reactivity: one bearing on the personality perspective and one on the social psychological view. The technical and procedural details will be kept brief, because the purpose of presenting these data is to examine a more general theoretical point.

CONTEXT SENSITIVITY OF REACTIVITY I: INDIVIDUAL DIFFERENCES

If our argument about the prevalence of the person-by-situation interaction is correct, then reactivity measured in one context should not be a potent predictor of reactivity measured in another. Consistent with this view, it has been quite difficult to find significant associations between laboratory stress testing responses and real world reactivity. Most studies in which generalizability has been examined have had subjects undergo reactivity testing in the laboratory and then wear an ambulatory monitor, which intermittently measures blood pressure and heart rate during subjects' normal activities. The laboratory change scores are then correlated with some measure of variability of the measurements taken in the field, such as the standard deviation. These studies have found, at best, small associations (24–34). One possible explanation is that there is an underlying measurement problem concerning the reliability and comparability of the measures being correlated: change scores and standard deviations. An alternative explanation, however, is that the change in the tasks performed, the setting, and the social context are enough to attenuate into virtual invisibility the cross-situational consistency of reactivity. An illustration of this is given in a study by Smith and O'Keeffe (35), who examined the cross-situational consistency of cardiovascular reactivity measures. They tested subjects twice, varying the setting, experimenter, measurement apparatus, and task between sessions. The correlations of reactivity responses from one setting to another were significant, but they were surprisingly modest. For systolic blood pressure changes, the correlation was .39, for diastolic changes it was .17, and for heart rate changes it was .43. Correlations of this magnitude leave more than 80% of the variance unexplained.

To explore further the generalizability of reactivity, we conducted a study examining the effects of context on cardiovascular responses. The reactivity of 24 subjects to mental arithmetic was tested on four occasions (36). We kept the task, the experimenter, and the apparatus constant. The testing was done with subjects in the same posture, at the same time of day, on four consecutive days. The only thing that we deliberately varied was the setting. Two of the sessions were conducted in the laboratory to provide a measure of the test–retest reliability. One session was conducted in a classroom, and one session was conducted in the subject's home to provide a measure of the extent to which simply changing the setting attenuates this consistency. Change scores were computed by subtracting subjects' resting blood pressure levels during a pretask baseline from their blood pressure levels assessed during the stressor. The means for these periods were computed using a pulse-based average (37). The stressor involved the subjects counting backwards by 13s with mild goading from the experimenter.

The data showed that resting blood pressure levels were quite similar across settings. Furthermore, the individual difference correlations assessed between different testing sites were no lower than the test–retest correlation. Thus, the resting blood pressures did not appear to be particularly sensitive to changes in context. This is consistent with the notion that resting blood pressure represents, in these data, a more purely physiological parameter, rather than a measure of the psychological state of the subject. It is also worth noting that blood pressure resting levels are a good predictor of future heart disease (38), though we will return to this point later.

Reactivity scores, at the level of group means, also show no effect of setting. On average, subjects' systolic blood pressure

TABLE 1
Pearson Correlations for Reactivity Scores Across the Four Settings ($N = 24$)

	Systolic Blood Pressure			Diastolic Blood Pressure			Heart Rate		
	Lab 1	Lab 2	Class-room	Lab 1	Lab 2	Class-room	Lab 1	Lab 2	Class-room
Lab 2	.68†			.62†			.09		
Classroom	.47*	.55†		.33	.26		−.13	.03	
Home	.44*	.32	.45*	.32	.25	.57†	.01	.23	.44*

Note: * $p < .05$, 2-tailed; † $p < .01$, 2-tailed.

increased about 14 mmHg, and diastolic increased about 4 mmHg from baseline. These reactivity scores at the level of individual differences, however, tell quite a different story. The correlations between the various settings for the two blood pressure measures are shown in Table 1. There was reasonable consistency in the scores between the two laboratory sessions for blood pressure, with correlations of .68 for systolic and .62 for diastolic. The heart rate change scores showed essentially no test–retest reliability, with a correlation of only .09. While the blood pressure correlations are not small, they do indicate that not even half of the variance in reactivity is shared across two sessions conducted in exactly the same way only one day apart. That is, there is a time-by-person interaction that accounts for more than half the variance. Viewed another way, a person's blood pressure reactivity to this task in this laboratory, with this experimenter, this equipment, and so on, is half determined by factors beyond our control and which vary from day to day. One can imagine any number of possible candidates: what subjects ate for breakfast, whether they had recently quarrelled with their roommates, how late they stayed up the night before, and so on. In our measurements, these things that vary randomly from day to day show up as noise. Conceptually, they indicate that reactivity is influenced by a number of things beyond what we can control in the laboratory, and so there is no number that fully captures a subject's reactivity even within our highly circumscribed setting. What happens, then, when we deliberately change some factor in our testing?

The average correlations between reactivity scores when the setting varied were .44 for systolic and .35 for diastolic. Simply moving from one setting to another was enough to drop the shared variance in half again. (Obviously, with a small sample, such percent-of-variance estimates are only approximate, but they are still useful as estimates of the effects of various alterations in the testing situation). These findings suggest there was a person-by-setting interaction accounting again for about half the variance in subjects' reactivity. This means that even if one could make the test–retest reliability perfect (by controlling everything that attenuated this correlation), there would still be only about half the variance shared between settings. For heart rate, because there was essentially no reliability to begin with, the effect of the change of setting cannot be assessed.

The change in setting is quite subtle. The laboratory is a smallish room, decorated in the usual medical warehouse style. The classroom is a larger space, filled with desk chairs. The apartments were suites in a residence hall, scattered with semi-dilapidated chairs and couches. If this change can influence the measures to such an extent, how great would be the influence of more substantive changes? How could one hope to predict the reactivity that comes from an angry interaction on the highway, or waiting for a root canal, or witnessing surgery, from the reactivity measured during mental arithmetic? That is, if blood pressure reactions to mental arithmetic in the laboratory are such poor predictors of blood pressure reactions to mental arithmetic that same week in a classroom in the same building, how well can one expect them to predict reactions to other stressors, much less heart disease, decades in the future?

CONTEXT SENSITIVITY OF REACTIVITY II: GROUP MEANS

We will now turn to the other approach to cardiovascular reactivity. To what extent can one use reactivity scores, averaged across people, as a measure of the situation? To address this issue, we will present data from another experiment. Unlike the previous study, the results are not presented elsewhere, and so we will provide a few more details. However, because the intent once again is to illustrate a general point rather than to test a specific hypothesis, we will not dwell on the minutiae of the procedure.

Ninety female undergraduates spent five minutes solving word search puzzles. These puzzles involved finding words embedded in a 10 × 10 grid of letters. Subjects were given one minute to work on each of five such puzzles, with the instructions to find as many words as possible. Two factors were manipulated—performance norms and encouragement. There were three levels of performance feedback. One-third of the people (30 subjects), in the feedback–success condition, were casually told by the experimenter at the beginning of the experiment that they should be able to find about three words in each puzzle. Pretesting had indicated that this was the standard number for such subjects. Another third were told, in the same offhand way, that they should be able to find seven words in each puzzle, a far higher number than they could reasonably achieve. This is the feedback–failure condition. The final third of the subjects, in the no-feedback condition, were not given any performance norms. Because of the nature of the task, subjects had no way of knowing how many words were embedded in the grid or, unless we told them, how well they should do at the task.

Orthogonal to the performance feedback manipulation, encouragement was manipulated by having the experimenter either pay attention to the subjects' performance or ignore it. In the encouragement condition (45 subjects), the experimenter gave encouragement on a predetermined schedule, periodically murmuring "nice one" or "good" after the subject found a word. This feedback was deliberately devoid of information about the subjects' performance relative to others' or any objective standard. For the other half of the subjects, the experimenter appeared not to be paying attention to task performance and did not provide any encouragement at all during the task.

We thus have a 3 × 2 design, with two levels of encouragement and three levels of performance feedback. This experiment allows us to address the question of how much reactivity is produced by the word search task and whether this depends on the quite subtle alterations in the situation. That is, is blood pressure response a measure of the task the subjects perform, or is it a measure of the context-specific social reality within which that task is performed?

Subjects' blood pressure was continuously monitored using an Ohmeda 2300 Finapres (for details about this instrument, see Wesseling [39]). A resting baseline level was assessed for 15 minutes, and subjects then spent 5 minutes working on a series of five word search puzzles. Reactivity scores were computed by subtracting the mean level during the final 5 minutes of the

baseline from the mean level during the 5 minutes of the task (again using pulse-based means [37]).

The pattern of blood pressure responses for each condition is shown in Table 2. Subjects who were not provided with any performance standard showed roughly half the reactivity of subjects who were given a standard, regardless of whether that standard was one they could match or not. A multivariate analysis of variance (MANOVA) indicated that this difference was significant for the blood pressure change measures, Wilk's Lambda = 89, multivariate F(approximation; 4,166) = 2.45, $p < .05$. Univariate analyses of variance (ANOVAs) indicated that the result was significant for both systolic and diastolic changes, with F(1,84) = 3.22 and 3.54, respectively, $ps < .05$. A post-hoc least significant difference (LSD) test indicated that for systolic change, the no-feedback condition changes were significantly smaller (at the .05 level) than the changes for either the feedback–success or feedback–failure conditions; and these two latter conditions did not differ from each other. The test showed that for diastolic changes, only the means for the no-feedback and feedback–success conditions differed from one another. The heart rate changes, consistent with the evidence for low reliability in the first study, reveal no significant patterns.

Table 2 also shows that the mean changes for the encouragement condition were substantially larger than those for the no-encouragement condition. A MANOVA indicated that the differences were reliable, with Wilk's Lambda = .92, multivariate F(approximation; 2,82) = 3.84, $p < .05$. Univariate ANOVAs indicated that the result was significant for systolic blood pressure changes, F(1,84) = 4.81, $p < .05$, although the result for diastolic blood pressure was not significant.

The results of this experiment demonstrate three things. One is that quite subtle changes in the context can have profound effects on the physiological response. All of the subjects did the same task, in the same room, for the same amount of time, and all did it in front of the same experimenter. The only alterations were the casual mention of performance norms before the task and the slight encouragement provided during the task, yet either of these was capable of doubling the response. It seems quite clear that the blood pressure reactions are not straightforward measures of the reactivity generated by a task, but rather tap the social context in which that task is performed. In other words, there is no set level of reactivity that is generated by a subject doing even this particular word search task. There is only a level of reactivity that is produced by this task in a specific context.

The second important finding of this experiment is that a manipulation which one might expect to influence the subjects' physiological responses had no effect. The subjects who were given performance norms which they could reasonably achieve showed essentially identical reactivity to subjects who were given a standard that guaranteed their failure. The simple mention of a standard influenced blood pressure, but the actual nature of that standard seemed, as far as blood pressure outcomes are concerned, not to matter.

The third relevant finding of this experiment is that the encouragement of the experimenter raised the blood pressure of the subjects. This seems like a perfectly reasonable result if one grants that such encouragement should increase the effort, interest, and excitement of the subjects. That is, such a finding can easily be reconciled with one's notions about how encouragement might work. However, notice that this exact manipulation could just as well have been called "social support." The encouragement condition could be considered the support condition, with the experimenter providing friendly comments and smiles, and the condition without encouragement could be the no-support condition with the absence of such positive regard. Conceptualized in this manner, however, the finding is just the opposite of what one would expect based on laboratory studies of social support (40–43). The supportive behaviors that have produced these effects have often been quite similar to the feedback of the present experiment: nods and smiles rather than indifference or inattention. In addition, the subjects of our experiment also reported that the encouraging experimenter was "more supportive" than the inattentive one. Using a 9-point Likert-type scale, subjects in the encouragement condition rated the experimenter as more supportive (M = 6.44, SD = 1.20) than did subjects in the no-encouragement condition (M = 3.67, SD = 2.21), and this difference was significant, with t(88) = 7.43, $p < .0001$. This could be regarded as a manipulation check, further supporting the notion that the results are backwards.

TABLE 2
Reactivity Scores for Word-Search Task ($N = 15$)

	Performance Norms		
	No-Feedback	Feedback–Success	Feedback–Failure
	Systolic Blood Pressure		
Encouragement			
No	3.1 (5.8)	8.6 (9.1)	9.1 (11.9)
Yes	7.8 (7.3)	13.2 (16.1)	14.0 (7.8)
	Diastolic Blood Pressure		
Encouragement			
No	3.8 (3.3)	7.9 (6.8)	5.5 (6.5)
Yes	4.4 (4.8)	7.8 (7.3)	8.0 (4.6)
	Heart Rate		
Encouragement			
No	5.1 (22.5)	6.2 (19.4)	15.6 (13.3)
Yes	5.6 (16.8)	9.8 (12.3)	8.9 (17.1)

Note: Standard deviations are given in parentheses.

Rather than debating whether we manipulated encouragement and produced the expected pattern, or manipulated social support and produced the opposite, we note that the effect of smiles and nods is highly interactive. In many contexts, such behavior has been shown to lower blood pressure, and in this context it raised it. In other words, this study, combined with social support effects of previous studies, is evidence for a manipulation-by-situation interaction. In one context, the presence of nods and smiles reduces reactivity, and in others it raises it.

One could argue that social support is, by definition, beneficial in terms of blood pressure response, and therefore, we did not create social support in this situation with our nodding and smiling audience. In other words, we do not necessarily have evidence for an underlying construct-by-situation interaction—it could be that an operationalization that in one context moves the underlying construct in one way will, in a subtly different context, move it quite another way. The difference between these two, however, is a subtle one and may best be left to those interested in philosophy and semantics. In any event, it is the case that one cannot count on reducing blood pressure responses to stress by adding a nodding and smiling audience.

There are any number of possible explanations for our finding that the smiling experimenter raised blood pressure responses. The task we used may not have naturally mobilized the full effort of our

subjects, and so encouragement could increase this effort. Our non-encouraging experimenter may also have been seen as non-evaluative, while in other designs the non-supportive audience might still have been seen as evaluative. Word-search ability may not be central to self-esteem, and so negative feedback may not constitute a threat. The exact nature of the feedback may determine whether the effect on reactivity is positive or negative. It could even be due to some factor invisible in standard methods sections, such as the distance between the audience and the subject or whether the audience was sitting or standing while delivering the feedback.

GENERAL DISCUSSION

The results of the two experiments suggest parallel conclusions. The data from the first suggest that one cannot get a pure measure of the relative reactivity of a particular person. Instead, the reactivity that an individual will show to a task depends on the context. The difficulty is in specifying what elements of the situation will interact with relevant person characteristics to affect the blood pressure response. The first study demonstrates that context can be as seemingly insignificant as the room in which the task is performed. That change is enough to dramatically attenuate the consistency of reactivity scores from session to session. The results of the second experiment suggest that one cannot get a pure measure of the reactivity associated with a particular task. Instead, the mean response of a group to a task depends on the context in which that task is performed. In the second study, the contextual variables that played a critical role in the blood pressure response concerned the experimenter's demeanor and a bit of information concerning the performance of other students at the same task, but not whether that information suggested subjects were failing or succeeding. Not only can subtle alterations in the situation have profound effects on blood pressure reactions, but it is difficult a priori to tell which manipulation will have an effect and whether it will heighten or dampen the cardiovascular response.

Given these findings about the context sensitivity of cardiovascular reactivity, what good are measures of blood pressure response? That is, how can one make any general statement, or generalize any finding, if physiological responses are so highly interactive? There are two possible answers to this dilemma. One is to ignore the interactions and hope that there is some general trend that is strong enough to be detected. The other is to match the research closely to real-world phenomena—to embrace psychological naturalism.

First, we will examine the approach of ignoring the interactions. In the case of the personality approach, while the person-by-situation interaction is large, there is some stability across settings or cross-situational consistency. It may be possible to predict later disease from reactivity measured in the laboratory, measured in the classroom, or measured in the home, even though these are all somewhat different measures. That is, one can view any response as some combination of that person's general reactivity (variance due to the individual), some context-dependent reactivity (variance due to the situation and person-by-situation interaction), and some noise. As long as the design is sufficiently powerful, the first—the person's characteristic reactivity—may provide some useful predictive ability. Researchers have, for example, been able to predict hypertension and increases in arterial plaque from relatively simple measures of laboratory cardiovascular reactivity (2–5, 10,11). Furthermore, it has been possible to estimate, from twin studies, the degree to which reactivity is inherited, again using laboratory reactivity designs (44). These findings suggest that measures of reactivity do capture something real and enduring about the individual and that success is possible even ignoring the interactions.

It is also possible to ignore the interactions in the social–psychological approach to reactivity. For the social–psychological approach, while one can shift the mean response of the group with subtle changes in the situation, it may generally be the case that one task produces more reactivity than another or that some intervention generally has some effect on the response. For example, smiles in most circumstances may lower blood pressure, and while assuming that this is a general effect may make false predictions in some contexts, it may be better than no prediction at all. Having a spouse may not be beneficial in all situations, but having a spouse appears to be helpful in enough situations that married people outlive their single counterparts (45).

The power of this approach can be enhanced by aggregating measures. Rather than trying to predict disease from one cold pressor response, one can predict from the mean of a dozen such responses. Similarly, rather than relying on one intervention to assess the impact of social support, one can average the effects of a dozen such interventions. Such aggregation has been suggested as a resolution for the more general cross-situational consistency problem in personality psychology (12). It is a basic psychometric principle that such aggregation will enhance reliability (an effect described by the Spearman–Brown formula). A single reactivity score may be analogous to a single item in a test battery. It may only have weak predictive power by itself; however, in combination with other items which tap the same underlying dimension, the item serves its purpose.

While aggregation can help cross-situational consistency to the extent that the measures are unreliable, aggregation cannot solve the person-by-situation interaction problem (46). In the data from our first study, for example, averaging over many responses to mental arithmetic would reduce the noise or person-by-time interaction. But even with an infinite number of measures, the finding that the predictability is cut in half by a person-by-situation interaction would not be changed. That is, even perfect measures of cardiovascular responses to the task in the laboratory would only predict a maximum of about half the variance in perfect measures of responses in the classroom or home. The same should hold for the social–psychological approach to reactivity. Assessing the effects of social support in the same manner on numerous occasions should yield a very precise indicator of the effect of that manipulation. However, as with the personality approach, maximizing reliability will not help much in estimating the effect of nods and smiles when they occur in a subtly different context. It will help predict what a different group of people would do in the same setting, but not what the same group would do in a different setting.

The obvious danger of ignoring the interactions and putting one's faith in the main effects is that predictive utility will be reduced and potentially important effects will be ignored. To the extent that person-by-situation interactions exist, reactivity measured in a particular context cannot reflect the person's true reactivity, let alone predict later disease. To the extent that manipulation-by-situation interactions exist, the effect of a manipulation in a particular setting will not be the same as its effect in another, let alone predict the effect of interventions on disease. It is then critically important not to assess reactivity or the effects of interventions in arbitrary settings, but rather to study them in settings carefully matched to the ones of interest. This is the second approach to the interactive nature of reactivity, one of psychological naturalism. For example, if one thinks that people who show

excessive reactivity in their daily lives are at increased risk of developing cardiovascular-related disease, then one should measure reactivity to daily life stress. If the stress that people experience—that causes frequent blood pressure and heart rate excursions—is from inserting one hand into ice water, then assessing reactivity to the cold pressor is reasonable. On the other hand, if the sort of stressor that people experience in the world is interpersonal, it would be wise to incorporate this into the reactivity testing. One essential element for this kind of work is paying careful attention to the phenomena of the real world. How often are people angry in the course of their lives? How does this compare to how often they are frustrated or anxious? Are these stressors generally interpersonal? Do they require action or simply endurance?

Creating this sort of psychological naturalism in the psychophysiological laboratory is not an easy matter. There is no one sort of stressor in the real world that is of interest, and so there is no one correct way to model stress in the laboratory. People experience anger, frustration, anxiety, and cold water in the course of their lives. Similarly, people receive approving smiles that they may sometimes interpret as support and sometimes as encouragement. This not only means that no one laboratory paradigm can capture all that is interesting about stress responses, but also suggests that field work will not avoid the problems of generality. Just because one has assessed reactivity to a real-world stressor in the field does not mean that the finding will be any more general than if one had studied a single laboratory response. However, studying a phenomenon in the laboratory that has been carefully matched to the real world or actually studying the phenomenon in the real world, while it will not guarantee generality, will at least ensure that the findings apply to at least one real thing.

The critical task is to determine just what real world phenomenon one wants to study, and then either study it as it occurs naturally or create that situation in the laboratory. There are, in any research design, an enormous number of parameters that have nothing directly to do with the issue under study but which still must be set at some level. Some of these are likely to be constant across sessions, such as how much the subject is told beforehand, the posture of the subject during testing, the size of the room, the visibility of the apparatus, and so on. Others may be allowed to vary randomly, such as the gender of the experimenter, how far the experimenter stands from the subject, and so on. Rather than picking these parameters arbitrarily or based on ease and convenience, they should be set to match the specific phenomenon of interest. It may well be that a person's response with these parameters set at one level will be quite different, in an interactive way, from that person's response with them set at another level. Likewise, the effect of a manipulation may vary depending on the way these details of the procedure are designed. The art of a good design is making these various decisions correctly—not so that some desired effect emerges, but so that the finding generalizes to the desired real world situation.

There are several clinically relevant examples of the importance of psychological naturalism. Many people who appear to be hypertensive when measured by a physician in a clinic setting exhibit normal levels when measured outside the clinic, even in ostensibly stressful settings (47). Such "White Coat Hypertension" has been estimated to account for approximately 20% of hypertension diagnoses (48). Measurement in the person's natural context captures something more important about the individual than the measurement in the clinic. It has also been found that resting blood pressure measured in the clinic does not predict left ventricular hypertrophy (LVH) as well as do ambulatory blood pressure measurements in the subject's natural environment during the subject's normal activities (49). These examples not only provide evidence for a person-by-situation interaction in resting levels (not just reactivity as we found), but also suggest that measuring the reactivity in the situation in which it might do damage adds predictive power. Recent work on the cardiovascular response to spousal debates is another example of work that acknowledges the importance of naturalism in the laboratory and incorporates the interactive nature of cardiovascular reactivity into the design (50,51).

Another implication of the psychological naturalism approach is that a different kind of aggregation may be possible. Rather than aggregating over repeated measures of the same phenomenon, one can aggregate over a variety of situations. For example, a person's reactivity score (or relative reactivity score) could be the average of responses to mental arithmetic, physical exercise, public speaking, and so on, ideally in roughly the same proportion as that found in daily life. Similarly, to assess the effects of social support, one could average the effects of interventions during mental arithmetic, exercise, speaking, and so forth. These results might be more representative of real world cardiovascular responses. There are a variety of research protocols that already are moving in this direction. For example, Kirschbaum, Pirke, and Hellhammer (52) use a protocol that assesses reactivity by averaging over a series of role-playing interactions. Kamarck et al. (53) aggregate over both tasks and sessions in order to reduce measurement error and enhance reliability.

However, there is some danger that such an approach will average over, and thus conceal, genuine and interesting phenomena. That is, the interactions are not just noise, but potentially important sources of information. Finding that social support lowers reactivity in some contexts and raises it in others does not just mean that the effect is unstable. It is also providing a hint that the phenomenon may best be understood at a different level. In other words, complexity is not necessarily inconsistency, but may instead reflect the failure to specify important moderating variables. This can be illustrated with an example from another domain. It has been found that an audience facilitates performance on easy tasks and impairs it on complex ones (23). This finding led Zajonc, in his social facilitation theory, to suggest that the phenomenon could be understood at the level of the dominant response, which is to do the task correctly when it is easy and incorrectly when it is hard. At this level, the phenomenon is not interactive: the effect of an audience is always to increase the emission of the dominant response. In such a case, it would clearly be a mistake to average over all tasks, easy and difficult alike, and try to abstract some global idea of what the general effect of an audience is on task performance. With social support, it could be that nods and smiles are not the most useful level of abstraction. It may be that anxiety is the critical variable, or effort, and social support does not always have the same effect on this construct. Only by attending to the interactions can one hope to understand the phenomena.

Given the evidence for the ubiquity of interactions across situations in cardiovascular reactivity, the appropriate construction of social reality is critical. There is no great trick to creating some social reality in the laboratory. People are always struggling to make sense of their situation and will always understand it in some way. There is no neutral situation in the laboratory and no cover story that will prevent subjects from forming some idea of what is expected of them or what the nature of their task will be. The way

that subjects understand their situation will influence their reactions, both overt behaviors and psychophysiological responses. The "psycho" cannot be taken out of psychophysiology.

REFERENCES

(1) Pickering TG, Gerin W: Cardiovascular reactivity in the laboratory and the role of behavioral factors in hypertension. *Annals of Behavioral Medicine.* 1990, *12:*3–16.

(2) Krantz DS, Manuck SB: Acute physiologic reactivity and risk of cardiovascular disease: A review and methodologic critique. *Psychological Bulletin.* 1984, *96:*435–464.

(3) Manuck SB: Cardiovascular reactivity in cardiovascular disease: "Once more unto the breach." *International Journal of Behavioral Medicine.* 1994, *1:*4–31.

(4) Matthews KA, Weiss SM, Detre T, et al: *Handbook of Stress, Reactivity, and Cardiovascular Disease.* New York: Wiley, 1986.

(5) Lovallo WR, Wilson MF: The role of cardiovascular reactivity in hypertension risk. In Turner JR, Sherwood A, Light KC (eds), *Individual Differences in Cardiovascular Response to Stress.* New York: Plenum, 1992, 165–186.

(6) Manuck SB, Kaplan JR, Clarkson TB: Behaviorally induced heart rate reactivity and atherosclerosis in cynomolgus monkeys. *Psychosomatic Medicine.* 1983, *45:*95–108.

(7) Manuck SB, Kaplan JR, Adams MR, Clarkson TB: Behaviorally elicited heart rate reactivity and atherosclerosis in female cynomolgus monkeys (Macaca fascicularis). *Psychosomatic Medicine.* 1989, *51:*306–318.

(8) Kaplan JR, Adams MR, Clarkson TB, Koritnik DR: Psychosocial influences on female 'protection' among cynomolgus macaques. *Atherosclerosis.* 1984, *53:*283–295.

(9) Kaplan JR, Manuck SB, Clarkson TB, Lusso FM, Taub DM: Social status, environment, and atherosclerosis in cynomolgus monkeys. *Arteriosclerosis.* 1982, *2:*359–368.

(10) Fredrikson M, Matthews KA: Cardiovascular responses to behavioral stress and hypertension: A meta-analytic review. *Annals of Behavioral Medicine.* 1990, *12:*30–39.

(11) Barnett PA, Spence JD, Manuck SB, Jennings JR: Psychological stress and the progression of carotid artery disease. *Journal of Hypertension.* 1997, *15:*49–55.

(12) Holroyd KA, Gorkin L: Young adults at risk for hypertension: Effects of family history and anger management in determining responses to interpersonal conflict. *Journal of Psychosomatic Research.* 1983, *27:*131–138.

(13) Houston BK: Personality characteristics, reactivity, and cardiovascular disease. In Turner JR, Sherwood A, Light KC (eds), *Individual Differences in Cardiovascular Response to Stress.* New York: Plenum, 1992, 103–123.

(14) Matthews KA: CHD and Type A behaviors: Update on and alternative to the Booth-Kewley and Friedmann quantitative review. *Psychological Bulletin.* 1988, *104:*373–380.

(15) Smith TW, Christensen AJ: Cardiovascular reactivity and interpersonal relations: Psychosomatic processes in social context. *Journal of Social and Clinical Psychology.* 1992, *112:*79–301.

(16) Hartshorne H, May MA: *Studies in the Nature of Character, Vol I: Studies in Deceit.* New York: Macmillan, 1928.

(17) Mischel W: *Personality and Assessment.* New York: John Wiley & Sons, 1968.

(18) Ross L, Nisbett RE: *The Person and the Situation: Perspectives of Social Psychology.* Philadelphia, PA: Temple University Press, 1991.

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

(20) Bem DJ, Allen A: On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review.* 1974, *81:*506–520.

(21) Funder DC, Colvin CR: Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology.* 1991, *60:*773–794.

(22) Festinger L: A theory of social comparison processes. *Human Relations.* 1954, *7:*117–140.

(23) Zajonc RB: Social facilitation. *Science.* 1965, *149:*269–274.

(24) Gerin W, Rosofsky M, Pieper C, Pickering TG: A test of laboratory-to-life generalizability using a controlled ambulatory procedure. *Psychosomatic Medicine.* 1994, *56:*360–368.

(25) Turner JR, Girdler SS, Sherwood A, Light KC: Laboratory-field generalization and intertask consistency. *Journal of Psychosomatic Research.* 1990, *34:*581–589.

(26) Harshfield GA, James GD, Schlussel Y, et al: Do laboratory tests of blood pressure reactivity predict blood pressure variability in real life? *Journal of Hypertension.* 1988, *1:*168–174.

(27) Ironson GH, Gellman MD, Sitzer SB, et al: Predicting home and work blood pressure measurements from resting baselines and laboratory reactivity in Black and White Americans. *Psychophysiology.* 1989, *26:*174–184.

(28) Langewitz W, Ruddel H, Schachinger H, Schmeider R: Standardized stress testing in the cardiovascular laboratory: Has it any bearing on ambulatory blood pressure values? *Journal of Hypertension.* 1989, *7:*S41–S48.

(29) Linden W, Con A: Laboratory reactivity models as predictors of ambulatory blood pressure and heart rate. *Journal of Psychosomatic Research.* 1994, *38:*217–228.

(30) Melville DI, Raftery EB: Blood pressure changes during acute mental stress in hypertensive patients using the Oxford intraarterial system. *Journal of Psychosomatic Research.* 1981, *24:*487–497.

(31) Van Egeren LF, Sparrow AW: Laboratory stress testing to assess real life cardiovascular reactivity. *Psychosomatic Medicine.* 1989, *5:*1–9.

(32) Parati G, Pomidossi G, Casadei R, et al: Limitations of laboratory stress-testing in the assessment of subjects' cardiovascular reactivity to stress. *Journal of Hypertension.* 1986, *4:*S51–S53.

(33) Floras JS, Hassan MO, Jones JV, Sleight P: Pressor responses to laboratory stresses and daytime blood pressure variability. *Journal of Hypertension.* 1991, *5:*715–719.

(34) Matthews KA, Manuck SB, Saab PG: Cardiovascular responses of adolescents during a naturally occurring stressor and their behavioral and psychophysiological predictors. *Psychophysiology.* 1986, *23:*198–209.

(35) Smith TW, O'Keeffe JL: Cross-situational consistency of cardiovascular reactivity. *Biological Psychology.* 1988, *27:*237–243.

(36) Gerin W, Christenfeld N, Pieper C: The application of generalizability theory to blood pressure resting levels and mental stress responses. *Journal of Blood Pressure Monitoring.* 1996, *1:*485–494.

(37) Glynn LM, Christenfeld N, Gerin W: Implications of alternative methods of computing blood pressure means. *Journal of Blood Pressure Monitoring.* 1997, *2:*175–178.

(38) Smith JJ, Kampine JP: *Circulatory Physiology.* London: Williams & Wilkins, 1990.

(39) Wesseling KH: Finapres, continuous noninvasive finger arterial pressure based on the method of Penaz. In Meyer-Sabellek W, Anlauf M, Gotzen R, Steinfeld L (eds), *Blood Pressure Measurement.* Darmstadt, Germany: Steinkopff Verlag, 1990.

(40) Kamarck TW, Manuck SB, Jennings JR: Social support reduces cardiovascular reactivity to psychological challenge: A laboratory model. *Psychosomatic Medicine.* 1990, *52:*42–58.

(41) Gerin W, Levy R, Pieper C, Pickering TG: Social support in social interaction: A moderator of cardiovascular reactivity. *Psychosomatic Medicine.* 1992, *54:*324–336.

(42) Lepore SJ, Allen KAM, Evens GW: Social support lowers cardiovascular reactivity to an acute stress. *Psychosomatic Medicine.* 1993, *55:*518–524.

(43) Christenfeld N, Gerin W, Linden W, et al: Social support effects on cardiovascular reactivity: Is a stranger as effective as a friend? *Psychosomatic Medicine.* 1997, *59:*388–398.

(44) Turner JR, Hewitt JK: Twin studies of cardiovascular response to psychological challenge: A review and suggested future directions. *Annals of Behavioral Medicine.* 1992, *14:*12–20.

(45) Berkman LF, Syme SL: Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda County residents. *American Journal of Epidemiology.* 1979, *109:*186–204.

(46) Mischel W, Peake PK: Beyond deja vu in the search for cross-situational consistency. *Psychological Review.* 1982, *89:*730–755.

(47) Pickering TG: *Ambulatory Monitoring and Blood Pressure Variability.* London: Science Press, 1991.

(48) Pickering TG, James GD, Boddie C, et al: How common is white coat hypertension? *Journal of the American Medical Association.* 1988, *259:*225–228.

(49) Devereaux RB, Pickering TG, Harshfield GA, et al: Left ventricular hypertrophy in patients with hypertension: Importance of blood pressure response to regularly occurring stress. *Circulation.* 1983, *68:*470–476.

(50) Smith TW, Brown PC: Cynical hostility, attempts to exert social control, and cardiovascular reactivity in married couples. *Journal of Behavioral Medicine.* 1991, *14:*581–592.

(51) Brown PC, Smith TW: Social influence, marriage, and the heart: Cardiovascular consequences of interpersonal control in husbands and wives. *Health Psychology.* 1992, *11:*88–96.

(52) Kirschbaum C, Pirke KM, Hellhammer DH: The 'Trier Social Stress Test'—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology.* 1993, *28:*76–81.

(53) Kamarck TW, Jennings JR, Debski TT, et al: Reliable measures of behaviorally-evoked cardiovascular reactivity from a PC-based test battery: Results from student and community samples. *Psychophysiology.* 1992, *29:*17–28.