

Policy Capturing Methodology in Sexuality Research

Michael W. Wiederman
Ball State University

Many research questions in sexuality involve assessing the stimuli that influence peoples' decisions and judgments. Typically, such research relies on self-reports of these influences, although it appears that humans have little or no insight into their own mental processes. Policy capturing is a more direct method for quantifying the factors influencing respondents' judgments, and this approach allows for assessment of the degree to which individuals have insight into their own cognitive policies. In this article, policy capturing methodology is reviewed, a recent application of the methodology is summarized as an illustration of the method as it may be used in sexuality research, and the strengths and weaknesses of the approach are discussed.

Although many research questions in sexual science have to do with behavior, some sexuality research is focused on people's judgments or decisions. In these cases, there is frequently an emphasis on *why* people make the choices that they do or the factors influencing those decisions. Examples might include studies of mate selection criteria, the conditions under which a condom would be used, decisions as to whether sexual activity would likely take place within a particular context, or the contextual factors influencing likelihood of engaging in extramarital sex. Note that although each of these examples involves behavior, the focus of study in each is *decision making*. Similar to studies focused on sexual behavior, the norm in these cases is to directly ask respondents about motives or the stimuli influencing their decisions, and the resulting responses are taken at face value.

Unfortunately, several problems exist with asking individuals to report on the factors that affect their own decisions and judgments. Researchers have convincingly shown that humans typically do not have good insight (or, many times, any insight at all) into the various influences involved in their decision-making processes, even though respondents typically believe that they do (e.g., Brehmer & Brehmer, 1988; Nisbett & Ross, 1980; Nisbett & Wilson, 1977). Despite objectively poor insight, people routinely report on the factors that went into their judgments or decisions. How can this be?

It appears that, when asked to comment on their own mental processes, which are humanly impossible to observe, respondents typically generate reports of motives and cognitive influences based on a sort of "folk psychology" as to what motives and factors are most plausibly at work. For example, self-reports on preferences and influences regarding intimate relationships may be tapping into individuals' relationship schemas (Baldwin, 1992), or

beliefs about relationship development and processes (Fletcher & Kininmonth, 1992), rather than actual influences on relationship decisions. In other words, what researchers end up measuring are people's *beliefs* about what influenced their judgment or decision. Sometimes these beliefs will coincidentally correspond to the actual influences, whereas other times they will not.

From a research standpoint, ideally the investigator would have the power to manipulate stimuli or conditions in people's lives and observe the resulting effects on their decisions and judgments. If I send a man with red hair to approach a particular female for sex, how does she respond? If I take the same male and change only his hair color (now blonde), how does the woman respond to his request? Assuming that only hair color changed from Trial 1 to Trial 2, any difference in the woman's response could be attributed to the man's hair color. Obviously, researchers do not have the ability to perform such experiments, especially when one wants to consider multiple variables and their possible interaction. Hence, researchers would more typically ask women, "To what extent does a man's hair color influence your decision to have sex with him?" Respondents will provide an answer, but the accuracy of those answers is dependent on, among other things, the degree of insight about the stimuli that influence their decision to have sex with particular men.

Is there a way researchers can possibly circumvent some of the inherent problems with introspection when studying the stimuli influencing people's sexual judgments and decisions? One can answer in the affirmative if one finds analogue laboratory conditions an acceptable research methodology. For example, given a specified judgment or decision-making task, one could manipulate or measure particular variables of interest and evaluate how variation in these variables is related to corresponding variation in the respondents' actual judgments or decisions. One such method for doing this is referred to as *policy capturing*.

POLICY CAPTURING

The term "policy" has come to be used in the field of

Address correspondence to Michael Wiederman, Ph.D., Department of Psychological Science, Ball State University, Muncie, IN. 47306-0520; e-mail: 00mwwiederma@bsu.edu

human judgment and decision making to refer to “the factors used in making a judgment and the relative weighting thereof” (Ullman & Doherty, 1984, p. 179). Within that context, the term “policy capturing” (PC) refers to “studies that analyze judgments made on the basis of multidimensional stimuli by means of a linear model” (Brehmer & Brehmer, 1988, p. 78).

Methodology

In a PC study the respondent is given a relatively large set of scenarios, each of which is composed of several stimuli, and the respondent is asked to make a judgment in response to each scenario (Stewart, 1988). The numeric values corresponding to each level of each stimulus (cue) within each scenario are then entered into a multiple regression equation to “predict” the respondent’s judgments. In this way, the relative importance of each cue in the respondent’s judgments can be quantified. The regression equation, with its indices of the relative weight given to each variable (e.g., Beta weights), represents the individual’s judgment policy, although alternative measures of the relative importance of the cues do exist (Stewart, 1988).

To be more specific, the analysis of individual policies involves treating each respondent’s judgments as a separate sample. This is why a relatively large number of scenarios must be presented to each respondent, as such a number must be large enough to permit multiple regression analysis of the individual’s judgments by regressing the judgments on the cues (Brehmer & Brehmer, 1988; Stewart, 1988).

For the sake of illustration, consider a task in which the respondent makes a series of judgments in response to 60 different scenarios, each of which consists of five cues (and each cue has several levels with corresponding numeric values). Let us represent the respondent’s judgment in response to any given scenario by Y_s . The 60 Y_s values are regressed on the 60 sets of five cues. The multiple correlation between the cues and the Y_s is a measure of how predictable (consistent) the respondent’s judgments are given the cues provided. Linear relationships, nonlinear relationships, and interactions among cues can be tested, as would be the case in other types of multiple regression analyses (Stewart, 1988).

Although this approach is inherently ideographic in nature, it is possible to aggregate individual policies to characterize the policies for a group of respondents (nomothetic approach), such as comparing men’s and women’s policies. However, the aggregation should occur only after the individual policies have been generated (see Hammond, McClelland, & Mumpower, 1980; Stewart, 1988).

How many scenarios should be used? Although there is not an answer set in stone (Brehmer & Brehmer, 1988), a general “rule of thumb” to ensure statistical stability is at least 10 scenarios per cue. So, if a researcher is asking participants to make judgments in response to scenarios comprised of five different stimuli, at least 50 scenarios should be used (see Stewart, 1988, p. 50). Greater numbers of scenarios should be used if the researcher (a) expects relative-

ly unreliable judgments by respondents, (b) uses cues that are intercorrelated, or (c) will be analyzing hypothesized interactions among cues. Frequently, a greater number of scenarios is used to allow for assessment of reliability (consistency) of respondents’ judgments by including some “duplicate” scenarios. That is, including a relatively small subset of duplicate scenarios embedded within the larger set allows one to correlate judgments across identical scenarios (sort of a test-retest reliability check).

Aside from including duplicate scenarios, the overall R value does give some indication of the reliability of the judgments. Multiple correlations of .7 to .9 are common in PC studies (Stewart, 1988), and a relatively high value (e.g., .8 or greater) is indicative of reliable judgments. A relatively low R value is open to several different interpretations.

It may be that the respondent was inconsistent in responding due to carelessness or a rushed approach to the task. However, if the stimuli presented to respondents for judgments contain information not controlled or measured by the experimenter (e.g., the use of photos as stimuli when not all characteristics of the photographed object have been quantified by the researcher), it is possible that the R value is relatively low because the respondent was basing his or her judgments on some quality not assessed by the researcher (and hence not included in the regression equation). This is less of a concern when the stimuli consist of written scenarios created by the researcher in which each aspect has been either measured (as one of the relevant cues) or held constant across scenarios. Another possible, though less likely, reason for a relatively low R value involves the respondent using a nonlinear configuration of the variables (cues) in arriving at judgments (Stewart, 1988).

An Example

Recent PC studies by Boon and Sulsky (1997), Finkelstein and Brannick (1997), and Wiederman and Dubois (1998) are examples of how the methodology can be applied to research topics having to do with sexually intimate relationships. I will consider the latter study in greater detail to illustrate use of the methodology in sexuality research.

Wiederman and Dubois (1998) were interested in testing evolutionary hypotheses regarding gender differences in preferred characteristics for short-term sexual partners. Specifically, would men and women place differential value on the characteristics of physical attractiveness, financial resources, generosity, prior sexual experience, current relationship status, and desired relationship status when evaluating potential short-term sexual partners? These authors constructed written descriptions of 50 hypothetical short-term sexual partners, each of which contained information about the six variables (cues) listed above. Each of these variables, or cues, had five levels, with a corresponding numeric value (1 to 5) associated with each level. Note that the numbers 1 to 5 were used to correspond with the ordinal nature of the five levels of each cue. However, the authors might have been more precise had they asked a group of the respondents’ peers to rate the

degree to which each level of the cue represented the construct. Then, those ratings could have been used as the values corresponding to each level of each particular cue.

To begin, Wiederman and Dubois constructed a "shell" for the scenarios, into which the corresponding level of the cue could be interjected. Below is the shell these authors used for male respondents (an analogous version was used for female respondents).

*When it comes to physical attractiveness, this woman is _____.
Financially speaking, she _____. Regardless of her financial situation, this woman _____. With regard to sexual experience, this woman has had sex _____. Right now she is _____.
Regardless of her current relationship status, this woman _____.*

After constructing five verbal descriptions for each cue, each corresponding to increasing levels or degrees of the characteristic, the authors randomly generated a value (1 to 5) for each cue for each description (scenario). In other words, scenario #37 might consist of a hypothetical short-term sexual partner who is described as having level 3 physical attractiveness, level 5 financial resources, level 1 generosity, and so forth. By randomly generating the level of each cue for each scenario, the cues were not significantly correlated, thus eliminating the potential problem of multicollinearity among the predictor variables (cues) in the regression analyses (Brehmer & Brehmer, 1988). Of course, sometimes a researcher might want the cues to be correlated to mirror real-world conditions (Brehmer & Brehmer, 1988; Stewart, 1988). Below is a single example of the 50 scenarios presented to respondents, in this case describing a potential female partner.

When it comes to physical attractiveness, this woman is above average; appealing to look at. Financially speaking, she is rather poor and does not have money for things beyond the bare necessities. Regardless of her financial situation, this woman is reluctant to share and does not frequently spend money on a dating partner. With regard to sexual experience, this woman has had sex with a total of 5 different partners and finds sex moderately enjoyable. Right now she is "going steady" with one man for whom she cares. Regardless of her current relationship status, this woman is looking for a potential spouse and hopes to get married before long.

College student respondents rated the desirability of each depicted short-term sexual partner using a 7-point scale. These ratings were then regressed onto the values for the 50 sets of six cues to derive a cognitive policy for each respondent. When the subsequent policies were aggregated across individuals, only two of the six hypothesized gender differences emerged with regard to PC preferences for short-term sexual partners. That is, men and women differed in the relative value placed on only two of the six cues when making judgments about the desirability of short-term sexual partners. However, when asked at the end of the task to rate how important the respondent believed each of the six characteristics were in determining his or her judgments, there were gender differences with regard to five out of the six.

Having data on both the PC value placed on each of the cues as well as the self-reported value placed on each of

the cues allowed Wiederman and Dubois (1998) to assess the degree of insight respondents had into their own cognitive policies. Interestingly, only with regard to physical attractiveness was there a statistically significant relationship between the PC and self-reported importance placed on the characteristic (cue). In other words, respondents who placed the most importance on the physical attractiveness of potential short-term sexual partners did report placing the greatest value on physical attractiveness, but otherwise respondents displayed a complete lack of insight overall into the partner characteristics that influenced their judged desirability.

Wiederman and Dubois (1998) concluded that, had they only used the typical self-report methodology, the evolutionary hypotheses would have been supported, as had been the case in previous studies using such methodology. However, the findings from the PC methodology resulted in different conclusions. It may be that direct questioning regarding the characteristics most valued in short-term mates results in responses based on gender stereotypes.

Strengths and Weaknesses

As mentioned above, the primary strength of PC methodology is the ability for the researcher to directly assess the relative effects of various stimuli (cues) on individual's judgments. Accordingly, the researcher is not relying on respondent insight regarding which stimuli were most impactful, but rather is measuring the impact each stimulus had on the respondent's judgments. Respondents can still be asked for such introspection, which would then allow the researcher to compare the individual's PC cognitive policy with that reported by the individual respondent (i.e., one can assess how much insight the individual has into his or her own policy).

That the methodology allows for ideographic analysis is another strength. Rather than simply assessing general differences between groups, one can examine each research participant's cognitive policy. So, true investigation of individual differences, and possible correlates of those individual differences, is possible (see Wiederman & Dubois, 1998). Rather than simply concluding that one group differs from another, variables that may determine who among the two groups is responsible for the apparent group difference can be examined.

Relying on introspection regarding motives for past behavior limits one's effective sample to those individuals who have engaged in the behavior. However, it may be of theoretical interest to compare the cognitive policies of experienced individuals with those who have not engaged in the behavior. An advantage of policy capturing is that one need not have experienced the depicted situation to be able to make a decision or judgment. For example, if one were studying potential influences in people's decisions to use a condom during vaginal intercourse, traditional methodology might result in asking respondents who have had coitus to recall the most recent experience and to report why or why not a condom was used. Policy capturing allows the

researcher to compare virgins and nonvirgins with regard to the impact of the variables contained in the scenarios, and any differences to emerge may have important implications for promoting condom use during first coitus.

An additional advantage of the PC methodology is that it allows the researcher a relatively high degree of control over the stimuli presented to respondents (especially when written scenarios are used). This strength is also one of the primary weaknesses of the methodology. It is possible that the stimuli presented to respondents are not representative of "real-world" stimuli encountered in the described situation, or that the researcher has failed to present the most relevant stimuli (Brehmer & Brehmer, 1988). If an important cue is not included in the study, its value in making judgments will go undiscovered.

To avoid such a possibility, researchers using a PC methodology frequently survey or interview the group from which respondents will be sampled first, to inquire about the most salient cues in the judgment task (Stewart, 1988). The researcher could also observe people actually engaged in making the judgments in their real lives in an attempt to determine the stimuli that are most important. Classic literature, mass media depictions, and clinical experience all provide possible sources of information for generating ideas about the most relevant factors involved in the particular judgment or decision-making task. So, a researcher interested in using policy capturing to assess the decision to end a marriage might first survey therapists who work with distressed couples, interview recently divorced individuals, and study accounts of divorce in magazines, television, and novels with regard to the factors that appear to be most important in the decision to divorce. The resulting stimuli could then be included in scenarios presented to respondents.

A related weakness of PC methodology involves the potential artificial nature of the stimuli presented to research participants. The stimuli may be *representative* of those encountered in the real world, but the *mode of presentation* may deviate significantly from that encountered in naturalistic settings (Brehmer & Brehmer, 1988). In this case, the judgment task is artificial and may result in judgments potentially not representative of those made in actual situations. How well do the stimuli presented to respondents, which typically involve a written description and/or photograph, compare to the actual sensory stimuli encountered in the "real-life" analogue? How well do "paper people" compare to real people as stimuli upon which to base a judgment? These are important questions and concerns to consider when conducting PC research (Brehmer & Brehmer, 1988).

All of this is not to say, however, that PC studies must involve written scenarios. For example, college students have reported in focus groups that extraneous stimuli such as the age, physical attractiveness, and dress of potential sex partners influenced whether the respondent judged that person a risk for HIV infection (Williams et al., 1992). One could test this possibility by having college students rate the physical attractiveness, perceived age, degree of provocative

dress, and so forth of a series of photographed individuals. A subsequent sample of college students then could rate each photographed individual with regard to perceived risk for HIV infection. Substantial correlations between the two ratings would support the idea that college students rely on particular extraneous cues for safety from HIV.

Another weakness of the methodology involves the very nature of the research task. Respondents are typically presented with 50, and sometimes upward of 100, scenarios, photos, or other depicted cases and asked to make a decision or judgment about each. The task can become tedious or boring, and maintaining respondent interest may be a concern. One possible solution is to break the task into more than one session (Stewart, 1988). I have had students express concern that using a PC methodology might result in respondents eventually "skimming" cases looking for the information that is most relevant to them rather than carefully reading each scenario or examining each photo. One can hope that research participants would be invested in the study and read every word or notice every detail; however, such is not always the case. When presented with standard questions regarding personal attitudes and behavior, hurried responding is liable to result in unreliable and potentially inaccurate responses. However, in responding to scenarios in a PC study, if the respondent has identified the one or two cues that are salient for determining his or her judgments, then focusing only on that information is not as problematic and is liable to result in reliable judgments across scenarios.

CONCLUSION

Policy capturing methodology is far from perfect, and there are several inherent weaknesses. Of course, it is applicable only when the focus of study is the *why* underlying judgment and decision making (or, more appropriately, the stimuli influencing judgments or decisions). When this is the case, it allows a direct assessment of the external factors that influence individuals' judgments, a quality that is particularly attractive given that humans apparently have little insight into their own mental processes (Gibbons, 1983). Also, based on research involving nonsexual judgments, it appears that there is much individual variation in what cues people use, how much insight each respondent has into his or her cognitive policy, and so forth. Policy capturing methodology allows for examination of these and other individual differences. Whether, and in what ways, it can be adapted to the myriad questions posed by sexuality researchers awaits further consideration.

REFERENCES

- Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin*, *112*, 461-484.
- Boon, S. D., & Sulsky, L. M. (1997). Attributions of blame and forgiveness in romantic relationships: A policy-capturing study. *Journal of Social Behavior and Personality*, *12*, 19-44.
- Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing? In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 75-114). New York: Elsevier.

- Finkelstein, M. A., & Brannick, M. T. (1997). Making decisions about sexual intercourse: Capturing college students' policies. *Basic and Applied Social Psychology, 19*, 101–120.
- Fletcher, G. J. O., & Kininmonth, L. A. (1992). Measuring relationship beliefs: An individual differences scale. *Journal of Research in Personality, 26*, 371–397.
- Gibbons, F. X. (1983). Self-attention and self-report: The "veridicality" hypothesis. *Journal of Personality, 51*, 517–542.
- Hammond, K. R., McClelland, G. H., & Mumpower, J. (1980). *Human judgment and decision making: Theories, methods, and procedures*. New York: Praeger.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.
- Stewart, T. R. (1988). Judgment analysis: Procedures. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 41–74). New York: Elsevier.
- Ullman, D. G., & Doherty, M. E. (1984). Two determinants of the diagnosis of hyperactivity: The child and clinician. In M. Wolraich & D. K. Routh (Eds.), *Advances in behavioral pediatrics*, Vol. 5 (pp. 167–219). Greenwich, CT: JAI Press.
- Wiederman, M. W., & Dubois, S. L. (1998). Evolution and sex differences in preferences for short-term mates: Results from a policy capturing study. *Evolution and Human Behavior, 19*, 153–170.
- Williams, S. S., Kimble, D. L., Covell, N. H., Weiss, L. H., Newton, K. J., Fisher, J. D., & Fisher, W. A. (1992). College students use implicit personality theory instead of safer sex. *Journal of Applied Social Psychology, 22*, 921–933.

Manuscript accepted July 13, 1998