A new approach to assessing observable attributes of the tax compliance behaviors of small business owners

John Bornmann, Rob Creekmore, Tobin Bergen-Hill

The MITRE Corporation, McLean, VA, 22012

Abstract. Measurement of individual behavior and attitudes are often based on the recall of respondents of actions that occurred in the past. These responses can often be biased due to incomplete or incorrect recall, or deliberate obfuscation by respondents. This paper discusses a new approach to data collection, through the use of a virtual reality environment, that allows researchers to collect data on respondents' tax decisions, at the moment they make them, through a simulation of a representative business year and subsequent tax preparation process. Disparate data can be collected and analyzed, providing researchers with greater reliability and validity in their research. A discussion of this data and potential implications on tax research is included.

Introduction

The IRS estimated a gross tax gap of $450 billion in 2006 (Tax Gap "Map", 2011), and it is likely that a significant portion of that gap was due to tax filing errors as well as tax avoidance and evasion[1]. Determining potential variables that lead to filing errors and tax evasion can help the IRS reexamine, experiment with, and improve its approach to taxpayer compliance, outreach, and education. MITRE has recently concluded the first phase of research into adapting virtual reality based serious game technology to allow the IRS to simulate taxpayer compliance behavior well enough to test hypotheses around taxpayer outreach intended to improve that behavior. By conducting a controlled experiment within a virtual reality based Serious Game, it should be possible to identify constellations of variables connected with filing errors and tax avoidance and evasion. Virtual games provide a mechanism to perform well-controlled experiments, modifying specific game play variables to determine the conditions under which tax filing errors, intentional and unintentional, are likely to occur.

Method

The virtual reality environment allows experimenters a wide range of options when it comes to conducting controlled experiments. In our test scenario, players were presented

[1] Tax avoidance in this paper refers to the legal and condoned techniques for reducing the amount of tax paid to the government, while tax evasion refers to reducing the amount of tax paid to the government through fraud, deception, and other illegal means.

with a sole proprietorship of an ice cream store. This scenario allowed experimenters to provide players with a situation which allowed for players to engage in a set of business decisions, such as inventory management, employee reimbursement, and quarterly tax preparation. This scenario was selected because the greatest single contributor to the gross tax gap was $122 billion from underreporting Business Income on Individual Income Taxes. Although sole proprietorships do not account for this complete amount, the processes and decisions made within this simulation should closely mimic the decisions leading to the tax gap.

Since the virtual reality environment allows external variables to be controlled, such as weather or inflation, experimenters can be more certain that actions are a result of experimental or measured conditions, and not environmental influences. In addition, other external variables can be introduced to measure the effects of different training or instruction sets. For the original pilot study conducted by MITRE, participants were divided into two cohorts and each cohort was provided with a different set of Schedule C instructions. The control group was provided a standard Schedule C instruction set, while the test group was provided a modified instruction set containing punitive language against taxpayers who provide false information. Although the sample size in the pilot study was too small to draw significant correlations between the test and control groups, given a larger sample size, it would be possible to determine if there were any differences in behavior based on differences in the instructions provided to players.

Another advantage of the virtual reality environment is the ability to track and maintain logs of player actions throughout the game environment. In this case, players were recorded through an external video camera that recorded their "real world" reactions to the game, as well as frame-grabbing software that captured their in-game activity. In-game chat logs and player decisions were also recorded to an external database system. Combined, these records of game play can allow the entire game experience to be analyzed as a collective whole, providing additional context for player decisions (Nacke, Lindley, & Stellmach, 2008).

Recording of game play may not be sufficient in and of itself to draw distinct and meaningful conclusions which would be helpful for IRS or other interested parties. In order to supplement this data collection, participants were also asked to complete an online survey, including questions about their in-game activity and a risk assessment index (Weber, Blais, & Betz, 2002), immediately following the game experience. The survey tool used for the pilot study is not intrinsically linked to the study itself, and could easily be replaced with other assessment tools or inventories, such as Kirchler and Wahl's TAX-I compliance inventory (2010). Our virtual reality technique could easily be used for validation of other tax compliance hypotheses, such as Torgler's theory of tax morale (2003) or Verboon and van Dijke's theory of sanction influence (2011).

In addition to quantitative measures, participants were asked a series of interview questions by game administrators, which were used to assess the game itself, as well as triangulate with the in-game and survey data for enhanced overall comprehension (Perlesz & Lindsay, 2003). Qualitative research can provide a fuller understanding of individuals' feelings about subjects. While a survey or other form of quantitative collection can provide seemingly verifiable information on beliefs, it can also be misleading. For example, in a study of perceptions of drug use, the beliefs of respondents were found to be more rich than a simple positive/negative dichotomy, but rather reflected diverse opinions not only about the moral judgments about drug use, but also of the role of government and society in policing the norms connected to those judgments (Buchanan, 1992).

Mixed methods research has a long tradition within the applied sciences, beginning with Norman Denzin's fundamental *The Research Act* (1978). In Denzin's formulation, triangulation can be either through multiple sources, or multiple approaches. It is generally agreed that multiple sources of data will increase knowledge and confidence of a topic, therefore it is easy to argue that multiple approaches (such as combining qualitative and quantitative methods) should improve these as well. Although mixed methods is traditionally seen as a combination of survey and interview data, our research identifies additional sources of data, collected virtually through the recording of game play, that can be combined with traditionally collected data and assist in both research and training within applied fields focusing on human behavior.

The post-game interviews thus provide researchers with the opportunity to gain a fuller understanding not just of the players' game choices, but their attitudes about those choices. For example, although keystrokes and reaction time to decisions can be measured and potentially categorize behavior, those categories are empty of meaning without the input of the players and what they were thinking during their decisions. Triangulation, combining both quantitative and qualitative data collection into a coherent whole, provides important context to illuminate in-game behavior.

Findings

Over the course of our pilot study, we collected a wide variety of data, from demographic information to game data to survey and interview responses. In order to conduct any useful analysis, we must first operationalize the various attributes that can be collected over the course of the virtual reality game experience. Although these attributes will never be able to be directly compared, without a proper understanding of the internal dynamics of each, and how it varies, analyses will continue to be shots in the dark. Here we describe the attribute, its source, the observation method, and recommended measures for each. The attributes which were collected in the pilot study are detailed in Table 1, below:

Table 1: Collected Attributes

Name	Source	Observation Method	Type	Format
Demographics	Player	Pre-Game Survey	Quantitative	Nominal/Continuous
Income	Game Script	In-Game	Quantitative	Continuous
Shrinkage	Game Script	In-Game	Quantitative	Continuous
Spoilage	Game Script	In-Game	Quantitative	Continuous
Non-player Character (NPC) Actions	Game Script	In-Game	Qualitative	Ordinal
Cheating	Game Play	In-Game	Quantitative	Ordinal
Risk Behavior Index	Player	Post-Game Survey	Quantitative	Ordinal
Interviews	Player	Post-Game Interview	Qualitative	Text

Attributes whose source is the Game Script are those which can be manipulated by the game designers in order to elicit different and distinct reactions from game players. Thus, the amount of income, shrinkage, and spoilage can be programmed directly into the game to determine the effects of variation in these on player behavior. These modifiable attributes are in addition to the change in the instruction set used in the pilot study, or any other external change.

Attributes whose source is Game Play are data collected in the traditional sense of the term: a measurement of the activities of the individual exposed to a specific stimulus. Although the pilot study only modified the Schedule C instruction set, it would be possible to vary any of the Game Script variables to observe possible changes in the game play of individuals and measure the effect of these changes on their tax behavior. For the purposes of the pilot study, the Cheating attribute was only measured as a bivariate (yes or no) proposition, but could potentially be scaled to reflect a level or recurrence of cheating behavior instead.

Player attributes are those which are theoretically independent of the game itself, including demographic variables such as age (continuous) or gender (nominal), and responses to the Risk Behavior Index, developed by Weber, Blais, & Betz (2002), which would indicate predilection for different types of risk, and which could be compared with cheating behavior within the game for a fuller understanding of the possible motivations

for cheating or fraud. In addition, the collection of true qualitative data, in the form of post-game interviews, allows for triangulation of the quantitative data with the underlying rationales and attitudes of players, improving our overall understanding of the game environment, the influences on players and, by extension, the public (Denzin, 1978).

Discussion

The virtual reality approach allows researchers to determine potential factors involved in tax filing errors as well as tax avoidance and evasion. However, we must clearly state that this research is not designed to profile or target individuals; we simply can use the environment to identify possible constellations of factors that could lead to changes in taxpayer behavior.

By recording the activities of players throughout the entire game, it is possible to isolate not only potential behavioral factors but also decision points during which a player spends more time in consideration of a specific choice. Player behavior can also potentially isolate thresholds for errors or cheating which can allow the development of improved taxpayer outreach, education, redesigned forms and instructions, and so forth. With post-game interviews, it is also possible to determine the underlying decision-making processes of players, and compare these with the activities within the game.

In addition, it is possible by using this sort of virtual reality approach to create a modular and adaptive research experience. The system itself is portable, either directly through delivering software to a possible player, or potentially opening up the virtual reality environment to remote users via web or other connections. This would allow the investigation of any number of additional factors, such as differences in region or other geographic elements, differences in levels of participation, or variations on a number of other variables. Of course, this portability would be at the expense of some qualitative data collection such as in person interviews or video recordings, but would increase the sample size and validity of experimental data.

These are not the only attributes which could be measured over the course of game play. In addition to inserting extra variables into the game- additional expenses such as utilities and property tax (not currently modeled within the environment), for example - it would also be possible to measure other attributes of players which the designers might consider relevant. For example, additional demographic information could be collected at the beginning or end of game play, or other indices could be included in the post-game survey, such as an alienation scale (Ray, 1982), or personal development measure (Constantinople, 1969).

Within the tax compliance literature, there are also a number of different scales and theories which can be assessed, such as those mentioned above. Kirchler and Wahl's TAX-I compliance inventory (2010) could be better validated through an experimental

approach such as we have developed, eliminating the problems of hypothetical scenarios with regard to questionnaire responses since participants in the virtual reality game are performing the behaviors directly rather than being asked to envision them. In addition, rather than asking participants to remember an action which occurred possibly months in the past (such as an annual tax filing), participants in a virtual reality game will have just completed the actions which they would be asked about, avoiding potential recall bias (Coughlin, 1990).

Conclusions

Although only a pilot study, the research conducted by MITRE exploring the use of virtual reality based serious games demonstrated the viability of this approach, and suggests ways in which virtual reality data can be used to improve our overall understanding of human behavior. Although limited to investigation into tax preparation and small business proprietorship, there are many other possibilities for the use of this environment. In the case of our pilot study, we collected data on demographics and attitudes among players, as well as recorded player activity through the course of the game to monitor for potential cheating or fraudulent behavior within the virtual reality environment.

The potential payoff of this research extends beyond simple data collection and triangulation as well. For example, the environment is not limited to data collection alone. It is possible to put players into different positions within the game, such as the role of tax preparer rather than proprietor, and use the environment as a training tool for practitioners. This sort of modification could also be used to ascertain thresholds for practitioner fraudulent behavior. While the current game puts the player in the role of proprietor, it could easily be recoded to allow a player to play a different role, such as a tax practitioner, and thus use the data collected to measure the practitioner's response to an NPC proprietor.For example, it would be possible to have the proprietor request that the tax preparer incorrectly file tax forms and to measure responses to that request. The virtual reality environment could also be used as a training tool for investigators to better understand potential intentional and unintentional mistakes among tax filers and practitioners. These are simply a few of the alternative uses for the sole-proprietor virtual reality environment currently developed.

The virtual reality environment also allows the game to be easily adapted to different arenas and environments, such as modeling different industries or business size. These modifications mean that this virtual reality environment scenario can be used by a variety of industries, including both private and government users, for training and evaluation, as well as data collection as discussed above. With simple modifications, the number of

employees, for example, can be increased, or even the number of storefronts. This substantially changes the game play, as the player will have a cascade of different choices to select from, as well as different pressures, and potentially different regulations, depending on the modeled environment. Although adaptation to different environments would require more effort on the part of a programmer to support a modification to scale, the current environment would allow for such modifications with little difficulty, providing a qualitatively different experience.

Overall, the virtual reality environment developed by MITRE provides a research, training, and evaluation tool, which has wide application. Through data collection both within and without the game, the tool allows researchers and evaluators to conduct controlled virtual reality based experiments. It can also serve as a training tool for tax preparers and other practitioners, or tax investigators, highlighting potential trouble spots within the current tax code and enforcement environment.

References

Buchanan, D. (1992). An uneasy alliance: combining qualitative and quantitative research methods. *Health Education Quarterly*, 19 (1), 117-135.

Constantinople, A. (1969). An Eriksonian measure of personality development in college students. *Developmental Psychology, 1*, 4, 357-372.

Coughlin, S. (1990). Recall bias in epidemiologic studies. *Journal of Clinical Epidemiology, 43, 1*, 87-91.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods.* New York: McGraw Hill.

Internal Revenue Service. (2011). *Tax Gap "Map" Tax Year 2006.* Retrieved from:
Kirchler, E., & Wahl, I. (2010). Tax compliance inventory TAX-I: Designing an inventory for surveys of tax compliance. *Journal of Economic Psychology, 31*, 331-346.

Nacke, L., Lindley, C., Stellmach, S. (2008). Log who's playing: Psychophysiological game analysis made easy through event logging. *Fun and Games: Second International Conference*. Eindhoven, Netherlands.

Perlesz, A. & Lindsay, J. (2003). Methodological triangulation in researching families: Making sense of dissonant data. *International Journal of Social Research Methodology, 6*, 25-40.

Ray, J. (1982). Toward a definitive alienation scale. *The Journal of Psychology: Interdisciplinary and Applied, 112*, 1, 67-70.

Torgler, B. (2003). To evade taxes or not to evade: that is the question. *Journal of Socio-Economics*, *32*, 283-302.

Verboon, P., & van Dijke, M. When do severe sanctions enhance compliance? The role of procedural fairness. *Journal of Economic Psychology*, *32*, 120-130.

Weber, E. U., Blais, A.-R., & Betz, N. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*, 263-290.