

## Chapter 2

# Massive Questionnaires

Traditionally, questionnaires designed by personality and social psychologists consisted of a few short batteries of related items. There were two reasons for this. First, using the old-fashioned pencil-and-paper technology, it was inconvenient to administer the questionnaires, and costly to enter the data into a computer by hand, so short questionnaires were especially desirable. Today, administering via computer eliminates the data entry task, and the respondents themselves may find answering more comfortable and thus be willing to answer more questions. Second, the goal of the research was to discover and verify one or two theoretical principles per study, each requiring more than one item to achieve reliable data, but best measured by a small number of items that had been laboriously pretested. We shall continue to value such well-designed measurement scales, but our purpose is very different, to capture the complex characteristics of a particular individual, which will require many new items as well as many well-established measurement scales.

During the 1980s, like many other sociologists I explored the possibilities for computer administration of questionnaires, but doing more of the programming myself than many of my colleagues chose to do, even publishing software that would allow students to construct their own questionnaires, administer them via computer, and then analyze the results using increasingly complex statistical analysis procedures [1]. Starting in 1997, I worked with other researchers to explore the potential not only for administering questionnaires online, but using the Internet to develop large numbers of questions that reflected the full range of popular opinions, rather than the theoretical predilections of scholars.

A transitional project was a series of studies carried out from 1974 until 1986 on public perceptions of space exploration, initially as a side study connected to my doctoral dissertation, *The Spaceflight Revolution*, that was a social history of the space program, and resulting in *Goals in Space*, a book about the diversity of viewpoints held by knowledgeable people about the potential benefits [2, 3]. The fundamental methodology was in two parts. First, I would include in one questionnaire a few open-ended questions asking respondents to write in their own words what they personally felt was a legitimate reason for supporting the

space program. Second, fixed-choice items for a second questionnaire would be distilled from all the text derived in the first questionnaire, this time asking respondents to rate the values of a range of well-defined ideas on a standard quantitative scale.

During the 1974–1986 period, it often proved necessary to write my own analysis software from scratch, for example writing a block-model clustering program as an alternative to the relatively limited factor analysis software then available, each run taking about 36 hours on an Apple II. This is extremely slow by today’s standards, but quite convenient given that I could be doing work on a different computer while that one dutifully processed data without the need for my supervision. The following chapter will bring that multitasking principle up to the present day, by considering how mobile and ubiquitous computing and communications can contribute to personality capture. The challenge here will be to build upon the introductory material in the previous chapter, about personality theory and questionnaire methods.

Tens of thousands of questions will be required to measure the full complexity of any individual’s personality, and the salience of any particular question will vary from person to person. Many competing personality theories exist, and our null hypothesis must be that all of them are true, but each applies only under certain conditions for certain people. Thus we need to develop flexible systems for gathering and collating information, using statistical tools like correlations and factor analysis, but not by any means limited to them. Indeed, one of the ways in which a person can be emulated is for another person to understand him or her, developing a mental model that represents that other person. Theory is not only a tool for organizing data and making decisions about what data to collect, but itself is a mode of emulation. To clarify these and many related issues, we shall begin with one of the classic measurement instruments, the MACH scale based on the personality of a single historical individual, Niccolò Machiavelli.

## 2.1 Machiavellianism

Niccolò Machiavelli (1469–1527) was among many things a very influential political theorist, and social science college students even today read his pair of short books, *The Prince* and *The Discourses* [4]. Machiavelli’s Wikipedia page suggests how his ideas are generally remembered: “He asserted that social benefits of stability and security could be achieved in the face of moral corruption. Aside from that, Machiavelli believed that public and private morality had to be understood as two different things in order to rule well. As a result, a ruler must be concerned not only with reputation, but also positively willing to act immorally at the right times. As a political scientist, Machiavelli emphasizes the occasional need for the methodical exercise of brute force, deceit, and so on.” [5]

Machiavelli’s works became very widely known; his memory was often reviled, and *Machiavellianism* came to be a term of opprobrium, signifying a pattern of

deceitful behavior. The 1970 book by Richard Christie and Florence Geis, *Studies in Machiavellianism*, reported the very promising results of a study to develop a questionnaire measurement scale, in which many items were based directly on Machiavelli's own words [6]. Subsequent research confirmed that this scale predicted some real-world duplicitous behavior, although it correlated to some degree with other measurement scales, and thus may not be an entirely distinctive characteristic. Perhaps it merely reflects honesty versus dishonesty. Imagine this simple, one-item measurement scale:

Think for a moment about your everyday interactions with other people. In general, how honest versus dishonest are you? Circle the one number below that best describes your typical degree of honesty:

Dishonest 0–1–2–3–4–5–6 Honest

Clearly, this question has problems. Perhaps an honest person will brood over it for many minutes, trying to remember various recent social interactions in which honesty was an issue, then finally circling the number 4, which represents modesty more than honesty. A dishonest person may answer quickly, circling the number 6, which claims to represent complete honesty, despite the fact the respondent was lying. This example illustrates more than just the thorny problem of how to elicit correct responses, because researchers in the heyday of traditional questionnaires developed a number of principles of scale construction that required multiple items to measure any important variable. Right away we see two: (1) the mind of an honest respondent is often better able to handle several questions about aspects of an issue, than a single question that lumps everything together, and (2) it is cognitively easier for the respondent to lie in response to one simple item than to a battery of items arranged in such a way as to elicit a complex pattern of responses. Both of these points recognize that question-answering is a cognitive task, that may be affected by the respondent's mental skills as well as the respondent's intentions. Often, such issues were discussed in terms of reliability and validity.

*Reliability* is the quality of an item or index that gives consistent results. A reliable item tends to get the same response from a respondent if administered twice, so long as the relevant circumstances have not changed. While some frequently-used measurement scales are single items, many are multi-item indexes because each response may include some random error, and combining several items tends to reduce their combined random error. Many statistical methods were developed to estimate the reliability of an index composed of several items, such as splitting the list of questions into two halves during the analysis, for example comparing the scores of the odd-numbered versus even-numbered items, or applying a summary statistic like Chronbach's alpha that is effectively the average of all possible split-half comparisons [7].

*Validity* is the quality of an item or index that measures the phenomenon it purports to measure. A valid item accurately reflects the desired aspect of the respondent's thoughts, behavior, or characteristics. A reliable index may be invalid, for example if it measures something different from what we assume it does. An unreliable index is probably not valid either. The validity of an item may vary depending upon the characteristics of the respondent, as the example of our honesty question

illustrates. We may use the term *conditional validity* to refer to situations in which the item itself is well-designed to be valid, but some respondents under some circumstances fail to understand it or otherwise fail to respond properly. This is why many of the questions in major public opinion polls are very simple, avoiding the use of words that some respondents will not recognize, and topics many people are not familiar with.

One way social psychologists attempted to deal with dishonesty was to develop special measurement instruments focusing on it directly but cleverly. Often this was conceptualized in terms of *yea-saying bias*, a tendency to agree with statements in a questionnaire quite apart from what they said, or *social desirability bias*, the tendency of a respondent to give socially acceptable answers to questions [8, 9]. Much effort was invested by several researchers to develop separate indexes of items to measure these forms of bias, but they encountered the problem that some people really did agree with the set of items in a yea-saying index, or so completely conform to social expectations that their natural responses were always nice and acceptable.

The historical period in which personology was developed, was one in which many psychologists believed that people were not fully conscious of their own thoughts and feelings, under the influence of the Psychoanalytic Movement. This may or may not be true, but it justified the arrogant belief on the part of some researchers that their respondents would naively answer all the questions in a questionnaire, blissfully ignorant of what the researcher was trying to accomplish. Much questionnaire research seems to find that educated people have very different attitudes from those of uneducated people, yet I always wonder when I see such a study whether the results came just from the fact that the more educated respondents could “psych out the study” and give socially desirable responses to the stupid researcher. With our focus on personality capture, we can hope that respondents will self-consciously develop a commitment to honest responding, and that with a very large number and wide diversity of measures we can capture the truth.

Reliability was not the only reason Christie and Geis developed Machiavellianism indexes composed of many items; they began with the goal to capture Machiavelli’s thinking more generally, and only later in a long research process distill its essence. The first step was to go through Machiavelli’s writings, in English translation, and copy out statements about human nature. Many had to be edited slightly to turn them into straightforward statements suitable for use in a questionnaire. At this point, the researchers felt they understood Machiavelli’s perspective, and of course he had written long ago in a different language, so they felt it was reasonable to add a few more similar items from their own experience of modern life.

Given that the work was being done before personal computers, the researchers used a typewriter to put each example on a separate  $3 \times 5$  file card. These were shown to colleagues, each of whom went through the set twice, first expressing personal agreement or disagreement with each statement, then explaining how they interpreted each one. Some statements proved to be ambiguous and were removed, leaving a set of 71 statements. These were assembled into a questionnaire administered to 1,196 college students. Each student was scored in terms of the 71 responses as high Machiavellian, medium, or low Machiavellian, and the two

extremes were compared. The 50 items that best distinguished respondents at the extremes became one major MACH scale, and then a subset of 20 became the most frequently used index, called MACH-IV. To control for the yea-saying bias, half were written so that Machiavelli would agree with them, half were rewritten so that he would disagree, and the sum of one group minus the other gave a single number as the Machiavellianism score.

A student of mine, Lyn Jacobson Hoefer, found this work very interesting, as did I, so we carried out a study to boil MACH-IV down further [10]. We administered a questionnaire containing the 20 items of the index to 810 college students, and then performed statistical analysis to see how the items related to each other. If MACH-IV were a perfect measurement instrument, we would expect the positively-phrased MACH items to correlate positively with each other, the negative-phrased one to correlate positively with each other, and items in different groups to correlate negatively. This proved mostly to be true, but some statements did not correlate strongly, and the strengths of the correlations varied even among the ones that did seem connected.

For example, one of the original statements extracted from Machiavelli's writings was, "Most men are cowards." I have some trouble seeing how that relates to being honest versus dishonest. Does a dishonest person lie to cover their cowardice? Or does dishonesty require a special kind of courage to brave the possibility of being unmasked? Perhaps cowardice and dishonesty are different facets of Machiavelli's negative view of humanity, but not closely related other than in their negativity. Christie and Geis actually reversed this item, to produce one of the 10 that Machiavelli would have disagreed with: "Most men are brave." That revision took the item one step further from Machiavelli's thinking, in an attempt to correct for yea-saying by the respondent. But the item has a different problem, namely that it refers to "men" rather than to "women" or to "people." To the male and female respondents of a decade later than when MACH-IV was created, this item may seem sexist, or just unclear. It did not correlate well in our data, so we dropped it out when we created a short form of the MACH scale.

Another item that did not work well had not come from Machiavelli's writings at all, but from popular American culture of the period before 1970: "Barnum was probably right when he said there's a sucker born every minute." This item might not make sense to respondents in later decades or outside the United States. "Barnum" refers to American circus entrepreneur Phineas T. Barnum (1810–1891) who frequently perpetrated hoaxes. Searching for the word "sucker" in Wikipedia takes one to the page titled "Gullibility," which defines that term as "a failure of social intelligence in which a person is easily tricked or manipulated into an ill-advised course of action." [11] This sucker item came from a different person and a different culture from the genuine Machiavelli items, but seemingly harmonized with his thinking because a sucker is the perfect victim for a liar. It did not correlate well with other items, for our respondents, so we dropped it out.

The result was a short-form 10-item MACH scale with an equal number of positively and negatively phrased items, which turned out to have two slightly different but mutually reinforcing subscales within it:

**High MACH****Tactics:**

Never tell anyone the real reason you did something unless it is useful to do so.

It is safest to assume that all people have a vicious streak and it will come out when they are given a chance.

It is wise to flatter important people.

**Perspective:**

Generally, people won't work hard unless they're forced to do so.

The biggest difference between most criminals and other people is that criminals are stupid enough to get caught.

**Low MACH****Tactics:**

Honesty is the best policy in all cases.

There is no excuse for lying to someone else.

One should take action only when sure it is morally right.

**Perspective:**

Most people are basically good and kind.

Most people who get ahead in the world lead clean, moral lives.

Machiavellians tend to agree with the High MACH items, and disagree with the Low Mach items. The tactics items advise a person about how to behave, while the perspective items describe the world the respondent inhabits. Thus an honest person living among dishonest people might score Low MACH on tactics but High MACH on perspective. For our respondents, the subscales correlated with each other at the 0.21 level, statistically significant but individual items tended to correlate higher within each subscale than across them.

After Hofer and I derived this short-form MACH index, I used it in a software-textbook educational package on questionnaire survey research, as part of a dataset students could analyze for practice. Respondents were 200 business executives, and the main focus was on popular management philosophies. Items were derived from management science publications, in the same way Christie and Geis derived items from Machiavelli's writings, and many of them functioned like extensions of the MACH scale. In some cases, the only clear connection to Machiavelli was the extent to which workers could be trusted to perform well, versus needing to be under strict control. For example, Table 2.1 shows 14 items derived from the pair of management philosophies contrasted in Douglas McGregor's classic writings [12, 13].

Douglas McGregor (1906–1964) was a professor of management, who proposed an influential framework that contrasted two different leadership theories, which he called X and Y. Theory X was more authoritarian and superficially looked somewhat Machiavellian, while theory Y was more trusting of subordinates and tried to motivate them through helping them achieve their own goals. Many readers assume, perhaps correctly, that McGregor himself believed in Theory Y and used Theory X merely to provide contrast and better advocate his own values. McGregor said that different managers out in the real world followed one or the other theory, although Theory X may have been in fashion early in the twentieth century, and Theory Y was more fashionable later on. This raises a general point that any person's conceptualization of a major issue will include alternative ways of thinking

**Table 2.1** Correlations between Machiavellianism scales and McGregorism items ( $N=200$ )

	Agree (%)	MACH	Tactics	Perspective
<i>Theory X:</i>				
X1. People are fundamentally lazy, irresponsible, and need constantly to be watched	5.5	0.18*	0.10	0.22*
X2. Most people must be coerced, controlled, directed, and threatened with punishment to get them to work hard for goals set by their employer	2.5	0.24*	0.14	0.29*
X3. The average human being prefers to be directed, wishes to avoid responsibility, has relatively little ambition, and wants security above all	17.0	0.24*	0.16	0.26*
X4. The average human being has an inherent dislike of work and will avoid it if he can	5.0	0.34*	0.17	0.44*
X5. To motivate his subordinates, a good manager will use the economic incentive of wage rises more than the intangible rewards of honor and respect	15.5	0.25*	0.18	0.24*
X6. The most important things a good manager does are to direct people's efforts, motivate them, control their actions, and modify their behavior to fit the needs of the organization	69.0	0.21*	0.17	0.16
X7. A good leader should give detailed and complete instructions to his subordinates, rather than merely giving them general directions and depending upon their initiative to work out details	31.0	0.00	0.02	-0.03
<i>Theory Y:</i>				
Y1. Under proper conditions, the average human being learns not only to accept but to seek responsibility	74.0	-0.03	0.08	-0.21*
Y2. People are fundamentally hardworking, responsible, and need only to be supported and encouraged	72.0	-0.32*	-0.19*	-0.37*
Y3. People can exercise much self-direction and self-control when their work satisfies their needs for personal achievement and social respect	96.5	-0.12	-0.04	-0.19*
Y4. Group goal-setting offers advantages that cannot be obtained by individual goal-setting	58.0	0.05	0.06	0.01
Y5. The most important thing a good manager does is to create a work environment in which people achieve their own personal goals best by working for the goals of the organization	89.5	-0.11	-0.07	-0.10
Y6. To motivate his subordinates, a good manager will use the intangible rewards of honor and respect more than the economic incentive of wage raises	51.5	-0.09	-0.02	-0.17
Y7. Most people have untapped resources of imagination, ingenuity, creativity, and other intellectual potentialities	86.0	-0.10	-0.01	-0.21*

\*Statistically significant beyond the 0.01 level

that the individual understands but rejects. This implies that personality capture must deal with both, documenting not only the person's favorite perspective, but also the competing perspectives the individual is able to conceptualize.

Thus, the 14 statements in Table 2.1 express McGregorism, just as the 10 items listed earlier express Machiavellianism, but the two subscales of McGregorism contradict rather than support each other. The 200 modern business executives seem to favor Theory Y over Theory X, with fewer than 20% agreeing with 5 of the 7 Theory X items, and majorities agreeing with all of the Theory Y items. However, they tend to agree with both X6 and Y5, which are central expressions of the two competing theories' management philosophies, or perhaps there is a grain of truth in both.

Three columns of Table 2.1 show the correlations between three versions of the MACH scale and agreeing with each of the 14 statements. For example, for the first statement the correlations are 0.18, 0.10, and 0.22. These are positive numbers, indicating that people who scored higher on Machiavellians were more likely to agree that people are lazy, compared with people who scores low on the scale. While correlations range from  $-1.00$  to  $+1.00$ , with questionnaire items that are phrased in different terms, the coefficients tend not to be very high. For 200 respondents, in data like this, correlations at or above 0.18 are considered statistically significant, because there is less than one chance in 100 that pure chance produced this result in the absence of any reliable connection between the variables.

The 0.18 is for the entire 10-item MACH scale; 0.10 is for the 6-item Tactics subscale, and 0.22 is for the 4-item Perspective subscale. Given that it has fewer items, the Perspective subscale measures less well than the two other scales, so the fact it has higher correlations than Tactics for six of the seven Theory X items indicates that Tactics are doing most of the work in connecting Machiavellianism with Theory X. In fact, the 200 modern managers score low on the overall MACH scale. The 10 MACH items were combined by adding the scores for the High-MACH items and subtracting the scores for the Low-MACH items. Each was rated on a five-point scale—strongly disagree, disagree, neutral, agree, and strongly agree—so the total scale could range from  $-20$  to  $+20$ . The mean score for these 200 respondents was  $-6.8$ , meaning they tended to be low in Machiavellianism, and indeed only eight of them had positive scores.

Item X7, about giving detailed instructions, seems not to fit with the other six Theory X items, and experienced managers may have found that detailed instructions really are necessary, regardless of what their management philosophy might be. The pattern of correlations between the MACH scales and the Theory Y items is complex, but there is some tendency for believers in the Machiavellian perspective to disagree with some of the Theory Y statements, or at least agree less strongly. Clearly, the 24 items involved in Table 2.1 have complex relationships to each other, in the minds of this particular set of respondents, but there is also evidence of underlying themes that connect the ideas. The table also connects the thought processes and thus the personalities of two deceased human beings, Machiavelli and McGregor.



The MACH scale illustrates how theory-driven research, based ultimately on the thinking of one individual human being, can produce useful questionnaire measurement indexes by boiling down many statements to a few. To produce massive questionnaires designed to measure the great complexity of an individual personality, we need to go in the opposite direction, from a few items to many, in the process combining the contributions of many people.

## 2.2 Ethnographic Questionnaires

There are many ways to combine scientific methods, and one that I used in the 1974–1986 spaceflight studies might be called *ethnographic questionnaires*. Ethnography, of course, is the documentation of a culture, usually conducted by cultural anthropologists using traditional qualitative observation techniques. Close reading of ethnographies written by highly influential anthropologists of the past shows they also made very great use of *native informants*, members of the society being studied who are able to articulate their culture especially well, whether via formal interviews or informal conversations. In the first phase of the ethnographic questionnaire method, the people answering the questionnaire serve as native informants to some degree, while the people answering items in the second questionnaire serve more like traditional respondents.

An example that highlights modern Internet-based communication technology and connects to the central theme of this book was a pilot study to explore human conceptions of the afterlife. In the previous chapter we saw how questions about the afterlife from the General Social Survey could be used to chart the conceptions held by members of an unusual subculture, and here we shall see how the items could be derived directly from the culture rather than from the minds and theories of social scientists. At the same time, this pilot study explored how web-based questionnaires could play a role in personality capture.

On May 23, 1997, I launched a website called The Question Factory, to prototype methods of online questionnaire development. It lasted about 2 years, using an Internet service provider named Erol's which was absorbed into a different company about the time I was shifting over to a team effort to be described below that carried out a pair of major online surveys garnering data from tens of thousands of respondents [14]. In November 1997, a Phase I questionnaire was added to the site, focusing on the afterlife. The first three open-ended questions clearly sought to learn the respondent's own, personal views, shown here with answers from one individual:

What do you BELIEVE will happen to you personally, after you die?

"Nothing really; maybe I might be reincarnated and my memory would be erased about my past life."

What do you HOPE will happen to you personally, after you die?

"I hope that I will live again sometime."

What do you FEAR will happen to you personally, after you die?

"Nothing."

Three other open-ended items asked the person to take the role of respondent, reporting what other people think, although of course filtered through this particular individual's own beliefs and perceptions.

Describe a belief that some people have about life after death, a belief with which you personally disagree.

"People say they will go to heaven and hell."

Describe a belief about life after death as you imagine it is held by people who belong to a very different culture and society from your own.

"Above."

Describe a belief about life after death that may have been held by people many centuries ago.

"Egyptians might have thought that Osiris, god of the dead, would let them have life in some afterlife."

Clearly, this person is aware of a range of culturally defined possibilities, and does not happen to adhere to the dominant Christian religion. The structure of these responses is typical for a thoughtful respondent, quite apart from their content, including the fact that some items stimulate answers that are full sentences, while other items stimulate only a perfunctory response. Thus, collecting a rich corpus of data does require multiple questions and multiple respondents, not only to obtain a diversity of responses, but to have a sufficient number of expressions of each particular idea in order to understand it fully. In the case of the Phase I afterlife questionnaire, 131 people submitted responses.

Qualitative methods sometimes called *grounded theory* were then used to produce a Phase II questionnaire comprised of 90 statements about the afterlife. Frankly, sloppy social scientists often use this term to describe a theoretical stumbling-around inside a culture, hoping that insights will somehow appear. However, as originally developed by Glaser and Strauss, grounded theory was a highly rigorous qualitative approach, that gradually developed a system of conceptual categories [15]. One part of the method was developing each category incrementally, by adding spoken or written text generated by members of the culture under study, until that category had become *saturated* and no longer changed as more text was added.

The Phase II questionnaire garnered data from 198 respondents, who rated each of the 90 statements on a 7-point scale, following this instruction: "How likely do you think it is that this will happen to you after you die?" In order to understand how online ethnographic questionnaires work, we shall examine the most detailed factor analysis that was done on these data, following the same general approach as every other factor analysis reported in this book, but identifying fully 18 factors, six of which were dominated by only a single item. With naturally derived data, such as from this pilot study, where the items were not selected because they were believed to represent theoretically significant ideas, it is common for early-numbered factors to comprise many items, and later-numbered factors to identify very minor dimensions of variation. However, if the goal is personality capture, the minor factors may really be hints of major conceptualizations that do not happen to be fully represented among the respondents, and thus the starting points for future research and instrument development.

<http://www.springer.com/978-1-4471-5603-1>

Personality Capture and Emulation

Bainbridge, W.S.

2014, VI, 213 p. 8 illus. in color. With online  
files/update., Hardcover

ISBN: 978-1-4471-5603-1