

### About CheatSweep

CheatSweep™ is IntelliSurvey's industry-leading proprietary system for removing potentially fraudulent and inattentive respondents from survey response data. We gather a variety of additional metrics while respondents are completing a survey, and store these metrics in our database along with regular response data. Then we automatically analyze that data and look for patterns that indicate possible cheating or inattention. Importantly, we also look for signs of good responses. For each factor, we assign a probability that this factor indicates cheating. For example, if a respondent fails an attention check, we might calculate that this factor, viewed independently, indicates a 95% chance that the respondent is cheating. On the other hand, if a respondent answers a long-form open-end with lots of valid words, we might decide that this factor indicates only a 10% chance that she is cheating. Then we combine these various probabilities using calculations derived from Bayes' Theorem. For example (without going into the math) if both of the above rules were true for a particular respondent, the combined probability of 95% and 10% would be 68%.

We apply this algorithm to all respondents who complete the survey, and thereby calculate an overall estimate of the cheating probability for each respondent. If this probability is over a pre-set threshold, we remove the respondent from the main sample and tag it with a status of "F".

This general algorithm is similar to the methodology used for spam detection. In spam detection, the probabilities depend on words. A word like "discount" might be more likely to be present in spam, so it would be assigned a probability above 0.50. A word like "erudite" on the other hand, might more often appear in legitimate email, and be assigned a probability below 0.50. (This per-word scoring is why spammers started putting invisible hidden words at the end of spam, which then triggered further work by spam detection software to filter out hidden text, and to consider the sequence of words.) In the end, spam software will combine multiple probabilities, determine a final value using Bayes' Theorem, and then allow or disallow the email based on this final score. CheatSweep is not so different from this, but it works using rules that leverage data collected during survey administration. The rules are based on IntelliSurvey's own analysis and experience in analyzing valid and cheating survey respondents.

### CheatSweep Processing and Thresholds

As described above, while a respondent is taking a survey, we gather a variety of extra data to detect behaviors that are relevant to cheating and attention. Then, when the respondent completes the survey, we do an analysis to determine the likelihood of cheating. This analysis can involve a fair bit of math and database number crunching, so we don't do it while the respondent is waiting. Instead, all completed responses are initially assigned a status of "C" for complete. Then the CheatSweep scoring process happens a short time later (usually within a few minutes). CheatSweep analyzes the data, comparing the response to other responses for certain rules, and to baseline values for other rules. If the final score for a response is above a certain threshold, it will then be moved to status "F" for "Possible fraud". In this status, the response will not count toward quotas, so we will naturally gather other responses to replace the potentially fraudulent response.

The threshold can be set in one of two ways:

A. By probability. That is, we can say "eliminate all respondents with an estimated probability of cheating above 90%".

B. By percentage of sample. With this method, we are saying "drop the worst X% of the sample." Here X could be 5% or 10% or any value we determine is appropriate based on the sample source and other factors.

In most cases, option A is preferable. That is because with option B, you will be dropping a certain percentage regardless of sample quality. For example, suppose you set the threshold at 5%. But if you have perfect sample, why drop even 5%? And on the other hand, if 25% of the sample is questionable, why should you keep the extra 20%? Generally speaking you want to keep the good and discard the bad, not assume a particular quality ratio in advance.

That said, option B may be appropriate in cases in which you know the sample quality with a high degree of confidence.

## Immediate Termination

As described above, CheatSweep applies a variety of rules to determine the likelihood of cheating, and normally runs after a respondent has completed the survey. However, there are two special cases that trigger immediate termination for respondents:

- Duplicate detection: if we are sure that a respondent has already completed the survey on the same device, we immediately terminate them without waiting for further data. (Note that this can be disabled for surveys taken from public locations, like a kiosk.)
- Wrong country: if the survey is configured to disallow respondents from particular countries, or to only allow respondents from certain countries, and we detect by the respondent's IP address that they don't match this criteria, we immediately terminate them.

## Frequently Asked Questions

The rest of this document provides questions and answers to common questions about CheatSweep. Please consult your IntelliSurvey project manager if you have additional questions that aren't covered here.

### How accurate is CheatSweep?

We believe that CheatSweep is the most robust and thorough survey cheating detection system available. However, it is important to remember that we can only make estimates of cheating probabilities. We unfortunately cannot look into a respondent's heart to discern his or her true intent. That being said, we can divide respondents into three broad categories:

1. Good: they seem to be obviously valid respondents. They pass attention checks, take the survey at a reasonable pace, answer open-ended questions, and otherwise seem to be answering honestly.
2. Bad: those that are obviously questionable. They might fail attention checks, go way too fast, straight-line tables, take the survey from different devices, and so on.
3. In between: these respondents might pass one attention check but go quickly, or perhaps take the survey from multiple IP addresses. There are some signs of possible problems, but other signs that they are valid.

Clearly, the third category is the most difficult to adjudicate. But because CheatSweep uses math and quantified analysis, we believe it will be better at categorizing this category than alternative methods. We don't throw somebody out just for straight-lining a table for example. They might answer that way for valid reasons, and have quality open-ended answers for example that

demonstrate attention and care. On the other hand, just passing a time check isn't enough to be categorized as "good", since inattentive respondents can slow down just to avoid disqualification. That's why CheatSweep uses a variety of rules, some of which influence the final score in a negative way, and some in a positive way.

But despite our best efforts and our contention that CheatSweep is better than alternatives, it is important to recognize that there are some gray areas. That is why the threshold value has an impact. And even with the best settings, it is likely that there will be some false positives (valid respondents identified as possible cheaters) and false negatives (cheaters that make it through).

### **What rules does CheatSweep use?**

CheatSweep uses approximately 20 different rules to determine the likelihood of cheating. This number changes over time, since we regularly add new rules and disable rules that are not adding any value. Some of these rules are based on standard industry tools, such as timing, and others are proprietary and unique to IntelliSurvey. Even the standard rules, like timing considerations, have been carefully adjusted so that they take into account things like survey branching, and so that they are not binary: extreme speeding is penalized more heavily than just somewhat above average pace.

### **Can I review the list of rules?**

IntelliSurvey does not publish the list of rules that are used for CheatSweep. There primary reasons are that some rules are proprietary, and releasing them would make it easier for potential cheaters to discover and work around them. In addition, we frequently create new rules and eliminate old rules that are not providing additional values, so the list of rules is not a static document that we can easily provide.

### **How do you create the rules?**

IntelliSurvey runs hundreds of survey projects per month, and we have gathered many millions of responses over the years. We use this data to continually calibrate CheatSweep. By looking for patterns in the data and applying our knowledge of survey design, we regularly design new rules, and then test them against our data, or try them on an experimental basis on live surveys. In addition as internet standards evolve, some rules are no longer useful and we can therefore eliminate them. As a result over time we attempt to continually improve the set of rules we use and thereby improve our cheating detection.

### **How important is it to use CheatSweep?**

We believe that it is critical, now more than ever. There are a variety of sample sources on the market, and it is sometimes difficult to determine sample quality. Some sample comes from pre-recruited panel, and other comes from the "river" such as social media posting or online ads. Regardless of the source, there are plenty of unethical respondents who are happy to lie or cheat to qualify for surveys and get whatever rewards are available. This problem is exacerbated by the fact that monetary rewards are more valuable, relatively speaking, in lower income countries that are often not part of the target sample.

## Can you implement my gate time rule or straight-lining rule?

Clients often come to us and ask us to implement speeding checks based on a particular time they have set. In our experience, this is a bad idea for several reasons. First, until the survey is online, it is difficult to get accurate timings. In addition, complex surveys have different branches, and so some respondents will reasonably complete the survey in far less time than other respondents, and it doesn't indicate cheating behavior. IntelliSurvey's algorithms automatically adjust for this, and so provide a much better way to detect cheating than a simple gate time. In fact, it may be that some respondents who read quickly and are adept at the subject matter complete the survey relatively quickly, but are still valid respondents. To throw them out can actually skew results and make the final data less reliable. For that reason, speeding checks should be a part of cheating detection, but not the whole story.

Similarly, using straight-line rules in an absolute way is a bad idea. IntelliSurvey's algorithms don't just check for straight-lining, but instead check for the variance of answers, and check multiple tables and other survey questions. This results in far better and more robust results than simple heuristics that our clients sometimes ask us to implement.

In summary, we have spent a lot of time building and calibrating CheatSweep, so using it as-is will be better than trying to add single-method rules to attempt to detect cheaters. For this reason, we discourage clients from providing us with their own rules that may conflict with our regular processing, and generally will provide less valid final results. We are however always open to suggestions, and we can even implement custom CheatSweep rules on a per-project basis if required. We can implement these in a way that fits into the overall CheatSweep framework rather than as a stand-alone metric.