

Oracle Heuristics

Version 6b Fall 2010

The lecture defines several oracle heuristics, based on work by James Bach and Michael Bolton (see Bolton's "Testing Without a Map", *Better Software*, January 2005, 24-28, at <http://www.developsense.com/articles/2005-01-TestingWithoutAMap.pdf>.)

In our experience, students need to try to apply these to testing situations before they understand them well enough to use them at work. This assignment gives you an opportunity to apply them in practice.

Review of the Concept

The point of an oracle is to help you decide whether a product's behavior is inappropriate--and if so, to help you explain persuasively to someone else why they should consider it inappropriate. In class, we studied oracle heuristics. The descriptive quotes below are from Bolton (2005), with minor editorial clarifications in [brackets].

- **Consistent with the vendor's image (reputation):** "The product's look and behavior should be consistent with an image that the development organization wants to project to its customers or to its internal users. A product that looks shoddy often *is* shoddy."
- **Consistent with its purpose:** "The behavior of a feature, function, or product should be consistent with its apparent purpose. [For example, help messages should be helpful.]"
- **Consistent with user's expectations:** "A feature or function should behave in a way that is consistent with our understanding of what users want, as well as with their reasonable expectations."
- **Consistent with the product's history:** "The feature's or function's current behavior should be consistent with its past behavior, assuming that there is no good reason for it to change. This heuristic is especially useful when testing a new version of an existing program."
- **Consistent within product:** "The behavior of a given function should be consistent with the behavior of comparable functions or functional patterns within the same product unless there is a specific reason for it not to be consistent."
- **Consistent with comparable products:** "We may be able to use other products as a rough, de facto standard against which our own can be compared."
- **Consistent with claims:** "The product should behave the way some document, artifact, or person [who has the authority to make promises about the product, such as a salesperson] says it should. The claim might be made in a specification, [a demonstration of the product], a Help file, an advertisement, an email message, [a sales pitch] or a hallway conversation."
- **Consistent with statutes, regulations, or binding specifications:** "The product [must comply] with legal requirements [and restrictions]." [The key difference between this oracle and *consistency with claims* is that the *claims* are statements made by the developer while statutes, regulations, and some types of specifications are imposed on the developer by outside organizations.]

Each oracle heuristic gives you a different type of inconsistency to consider.

If the program operates inconsistently with one of the heuristics, the structure of your argument is, "This [aspect of the program] appears to be unacceptable because it is inconsistent with the way we believe it should work. The basis for our belief is <history of the product><specification><etc.>"

Applying the Concept

When people are learning to use these heuristics, they often face similar challenges. Here are six questions that you should try to answer for the heuristics that you work with in this assignment

1. What is this heuristic? Explained in your own words, what does it mean for something to be inconsistent under this heuristic?
2. Will your primary research for this heuristic come before or after you run the tests?
3. What sources of information should you research and how will you find them?
4. What counts as an "inconsistency"?
5. Is this particular inconsistency bad, good or neutral?
6. Can we determine whether a particular inconsistency is minor or serious?

The Assignment

We'll do this assignment in phases.

- In the first phase, the instructor will assign you to a discussion group. You'll work through the problem together.
- In the second phase, you should work with your group to answer 6 questions.
- In the third phase, the instructor will assign you to a problem that you should work on your own.

Phase 1: Consistency with Our Image

Once you understand a company's image, you can assess the significance of a problem in terms of that image. The question that you are asking is, will this problem surprise people or cause people to lose respect for the company or lose their trust in the company?

Consider three manufacturers: Apple Computer, Microsoft, and Blizzard. If you don't know their products, look them up **quickly** on the web. Don't spend more than 30 minutes per company. (From watching students working on this problem in university classes, we think most students can find enough information in 5 minutes.) These three companies have quite different reputations. What are the differences?

It does not matter, for this exercise, if your impressions are absolutely correct. Do the best you can, as a group, in the limited time you have been assigned.

Post the answers in the discussion forum assigned to your group.

- 1) For each company, make a list of three attributes—three key words that describe the company's image (its public reputation or the public reputation of its products).
- 2) All three of these companies have made computer games, so imagine a computer game that has problems.

- a) Imagine first a problem that you would consider inconsistent with Apple's image, but not a surprise for (not very inconsistent with the image of) Microsoft or Blizzard (or both). Describe the problem and why you think it is consistent or inconsistent with the image of each company.
- b) Now imagine a problem that would be inconsistent for Microsoft but OK for Apple or Blizzard (or both). Describe the problem and why you think it is consistent or inconsistent with the image of each company.
- c) Finally, describe a problem that would be inconsistent for Blizzard but not for Microsoft or Apple (or both) and explain why you think it is consistent or inconsistent with the image of each company.

Please try to create a group answer. If necessary, include up to 5 attributes (so that everyone has their say). It is NOT essential that you all agree on the 3 "correct" reputational descriptors for each company. You haven't had enough time to do perfect research. What IS essential is that, after you have decided what some descriptors ARE for a company, you then figure out what attributes or behaviors of the program would be inconsistent with each those descriptors.

Phase 2: Consistency with Our Image

Do this task with your group. Now try to answer the following questions (the "Applying the Concept" questions) as if you were testing a computer game made by Blizzard.

As part of your process of developing answers to these questions, please tour the answers provided by the other groups to discover new approaches to some of these tasks.

- Please post feedback on other students' work as appropriate, and in your answers here,
- Please give credit to ideas you gathered from other groups.

1) WHAT IS THIS HEURISTIC?

What does it mean for something to be consistent or inconsistent with a company's image? For example, if a company has a reputation for making junk, is a junky product inconsistent with the company's image? Why or why not?

2) RESEARCH

- i) What research might you do *before testing*, to help with your ability to apply the consistent-with-image heuristic?
- ii) Why would you do research before testing (what are you trying to learn?)
- iii) What research might you do *after some testing*, after finding some results (e.g. some suspicious product behavior), to help you apply the consistent-with-image heuristic?
- iv) Why are you doing post-test research? What types of things are you trying to learn?
- v) Do you think you would do more consistency-with-image research before testing or after? Why?

3) RESEARCH SOURCES

Apart from looking at Blizzard's website, how could you learn about Blizzard's image/reputation? List your group's best three ideas.

4) WHAT COUNTS, UNDER THE CONSISTENCY-WITH-IMAGE HEURISTIC, AS AN INCONSISTENCY?

Would it be possible to find a bug that is neither consistent, nor inconsistent, with the company's reputation or public image? Why do you think so? (The best answer might have an example or two.)

- 5) HOW WOULD YOU DECIDE, UNDER THE CONSISTENCY-WITH-IMAGE HEURISTIC, WHETHER A PARTICULAR INCONSISTENCY IS GOOD, BAD, OR NEUTRAL?

HINT: Could there be an inconsistency with your company's reputation that is not a bug?

- 6) HOW WOULD YOU DECIDE, UNDER THE CONSISTENCY-WITH-IMAGE HEURISTIC, WHETHER A PARTICULAR INCONSISTENCY IS MINOR OR SERIOUS?

HINT: If a serious bug wouldn't negatively change your company's reputation, would it be serious under this heuristic?

Phase 3. Apply this to another heuristic.

You will be assigned to a forum that focuses on one oracle heuristic. The goal is to do a heuristic analysis yourself.

Discussion 1: Consistent With Purpose

Please answer these questions in the Consistent-With-Purpose forum. Answer the questions yourself first, then look at the answers of the other students and post your feedback on their answers. If you wait until the last minute to post your answers, you won't get any feedback, you won't have time to give any feedback, and you will learn less from this assignment.

The lecture presented an example of a bug that would be inconsistent with Microsoft Word's purpose but not with WordPad's purpose. In this case, the inconsistency was with the broad purpose of the entire program:

- Word is a professional-quality word processing program that promises precise control over formatting
- WordPad is a free utility editor that displays formatted text, and has limited formatting capability. It offers a convenient way to look at documents and create simple documents but it is not a production-quality word processor.

More often, under this heuristic, we look at a bug in a feature and ask whether the bug is fundamentally inconsistent with the purpose of that feature. Bolton's article gave the example of an unhelpful help message.

- 1) Imagine inserting and formatting a table in a Microsoft Word document (or, if you prefer, Open Office Writer). Now imagine two bugs:
 - a) The first bug should be inconsistent with the purpose of inserting and formatting a table in a Word document. Describe (briefly) that bug. Feel free to make up a bug. It doesn't have to be a real bug. Explain why it is inconsistent with the feature.

NOTE: Please don't give us a bug like, "IT CRASHES WHEN I INSERT A TABLE." Give us something more interesting than this-feature-doesn't-work-at-all.
 - b) The second bug should be a bug associated with inserting or formatting a table (you can make this bug up too). It should be a misbehavior that any reasonable person would agree is a bug, but it should not be inconsistent with the purpose of this feature. Explain why it is not inconsistent with the feature.

- 2) Now please answer the 6 basic questions, for this heuristic:
 - a) What is this heuristic? Explained in your own words, what does it mean for something to be inconsistent under this heuristic?
 - b) Will your primary research for this heuristic come before or after you run the tests?
 - c) What sources of information should you research and how will you find them?
 - d) What counts as an “inconsistency”?
 - e) Is this particular inconsistency bad, good or neutral?
 - f) Can we determine whether a particular inconsistency is minor or serious?
- 3) Now please give feedback to the other students who answered these questions.
- 4) If you have time, please look over the posts in at least one other discussion (analysis of one other heuristic) and give feedback / ask questions in that discussion as well.

Discussion 2: Consistent With User Expectations

Please answer these questions in the Consistent-With-Purpose forum. Answer the questions yourself first, then look at the answers of the other students and post your feedback on their answers. If you wait until the last minute to post your answers, you won't get any feedback, you won't have time to give any feedback, and you will learn less from this assignment.

What is the difference between relying on user expectations and relying on specifications? We know that if there are specifications, they can influence user expectations, but what other sources of information do users rely on?

- 1) Imagine two products, one that is likely to be very thoroughly specified (e.g. a medical device) and another that is not so thoroughly specified (maybe a computer game or a built-in-a-hurry web app). Be specific--choose two products.

In both cases, imagine sources of information that would tell you about user expectations. Brainstorm a long list, then enter your best 5 ideas for each product.

For each product, imagine two bugs (or other problems with the product). You can make these up. They don't have to be real bugs, but they have to be good examples.

- a) For the thoroughly-specified product:

- The first bug should be inconsistent with user expectations. What is this bug? Why do you think it is inconsistent with user expectations? What evidence supports your judgment? (If you don't have evidence, what kind of evidence would you look for, that COULD support your judgment?)
- The second bug should **not** be inconsistent with user expectations. Describe this bug. Why do you think it is consistent with user expectations? What evidence supports your judgment? (If you don't have evidence, what kind of evidence would you look for, that COULD support your judgment?)

- b) For the lightly-specified product:

- The first bug should be inconsistent with user expectations. What is this bug? Why do you think it is inconsistent with user expectations? What evidence supports your judgment? (If

you don't have evidence, what kind of evidence would you look for, that COULD support your judgment?)

- The second bug should **not** be inconsistent with user expectations. Describe this bug. Why do you think it is consistent with user expectations? What evidence supports your judgment? (If you don't have evidence, what kind of evidence would you look for, that COULD support your judgment?)

2) Now please answer the 6 basic questions, for this heuristic:

- a) What is this heuristic? Explained in your own words, what does it mean for something to be inconsistent under this heuristic?
- b) Will your primary research for this heuristic come before or after you run the tests?
- c) What sources of information should you research and how will you find them?
- d) What counts as an "inconsistency"?
- e) Is this particular inconsistency bad, good or neutral?
- f) Can we determine whether a particular inconsistency is minor or serious?

3) Now please give feedback to the other students who answered these questions.

4) If you have time, please look over the posts in at least one other discussion (analysis of one other heuristic) and give feedback / ask questions in that discussion as well.

Discussion 3: Consistent With History

Testing for consistency over the history of a product is not just about testing for exactly the same behavior over time. When we look for consistency, we look for similarities. When we see a case in which the program behaved THIS way with THIS feature, we try to imagine the different predictions we could make about how it will behave with other features.

Usually, we get interested in consistency with history for one of two reasons:

- We think we are seeing a new kind of weakness in the programming of the product.
- We think we are seeing a user interface change that will dissatisfy people who have worked with this product in the past.

Note that if the product has had essentially the same problem for many years, that will probably change how the development prioritizes fixing the problem. If you know that the problem has been around for years, it is important to acknowledge this, so the programmers don't get fooled into thinking they are working on something new.

OpenOffice is an example of a product with a long history. There have been many versions released over many years. One of the programs in OpenOffice is the spreadsheet.

Suppose you were testing the spreadsheet and you discovered that when you attempted to correct a very long word, OpenOffice's spellchecker would replace the old word with one that is always shorter than the original, even if this is not what you specified.

1) How would you determine whether this is consistent with the history of OpenOffice? What sources would you check? Would you limit your investigation to the spreadsheet? Why or why not? If not, where else would you look? Would you only look for spellchecker misbehavior that replaces long

words with short words or would you look for other ways that the spellchecker has corrupted the document? Why?

- 2) Now please answer the 6 basic questions, for this heuristic:
 - a) What is this heuristic? Explained in your own words, what does it mean for something to be inconsistent under this heuristic?
 - b) Will your primary research for this heuristic come before or after you run the tests?
 - c) What sources of information should you research and how will you find them?
 - d) What counts as an “inconsistency”?
 - e) Is this particular inconsistency bad, good or neutral?
 - f) Can we determine whether a particular inconsistency is minor or serious?
- 3) Now please give feedback to the other students who answered these questions.
- 4) If you have time, please look over the posts in at least one other discussion (analysis of one other heuristic) and give feedback / ask questions in that discussion as well.

Discussion 4: Consistent Within Product

Testing for consistency within a product is not just about testing for exactly the same behavior in an exactly comparable part of the program. When we look for consistency, we look for similarities. When we see a case in which the program behaved THIS way with THIS feature, we try to imagine the different predictions we could make about how it will behave with other features.

Usually, we get interested in consistency within-product when we think we are seeing a user interface oddity. Users will expect this to work the same way as the rest of the product and the inconsistency will confuse and dissatisfy users.

We often do a within-product consistency check for programming errors, but that’s not for oracle purposes. In this case, we’re not trying to decide whether X is a bug or not. We’re trying to investigate how much of the program a type of bug has infested. The programming group will often prioritize more highly (try harder to fix and fix sooner) a bug that impacts many parts of the program and will prioritize as less urgent a bug that only appears in one small corner of the program.

OpenOffice is an example of a product with a many components that are closely related and share a common look and feel. (If you prefer, you can think in terms of Microsoft Office.) One of the programs in OpenOffice is OpenOffice Calc, the spreadsheet.

Suppose you were testing Calc, and you pasted a picture into a cell of the spreadsheet. Then, when you tried to copy and paste the spreadsheet row or column that held the cell with the picture, the text and numbers copied and pasted, but not the picture. Suppose you find this surprising and don’t like it.

- 1) How would you determine whether this is consistent within the OpenOffice suite of products? What sources would you check? Would you limit your investigation to the spreadsheet? Why or why not? If not, where else would you test or do research?
- 2) Now please answer the 6 basic questions, for this heuristic:
 - a) What is this heuristic? Explained in your own words, what does it mean for something to be inconsistent under this heuristic?
 - b) Will your primary research for this heuristic come before or after you run the tests?

- c) What sources of information should you research and how will you find them?
 - d) What counts as an “inconsistency”?
 - e) Is this particular inconsistency bad, good or neutral?
 - f) Can we determine whether a particular inconsistency is minor or serious?
- 3) Now please give feedback to the other students who answered these questions.
 - 4) If you have time, please look over the posts in at least one other discussion (analysis of one other heuristic) and give feedback / ask questions in that discussion as well.

Discussion 5: Consistent With Comparable Products

Testing for consistency within a product is not just about testing for exactly the same behavior in an exactly comparable part of the program. When we look for consistency, we look for similarities. When we see a case in which the program behaved THIS way with THIS feature, we try to imagine the different predictions we could make about how it will behave with other features.

Usually, we get interested in consistency with comparable products when we think:

- Our user interface is strange in some way and users who have experience with other products will be confused and dissatisfied with our design.
- Our product is more unreliable (fails more often, or worse) in some annoying way than the competition.
- Our product has a bug that makes us think the programmer just doesn’t understand a specific task and we want to be able to say, “Everyone else does it **this** way instead.”

Suppose you were testing Calc, and you pasted a picture into a cell of the spreadsheet. Then, when you tried to copy and paste the spreadsheet row or column that held the cell with the picture, the text and numbers copied and pasted, but not the picture. Suppose you find this surprising and don’t like it.

- 1) How would you determine whether this is consistent with comparable products? What sources would you check? Would you limit your investigation to spreadsheets? Why or why not? If not, where else would you test or do research?
- 2) Now please answer the 6 basic questions, for this heuristic:
 - a) What is this heuristic? Explained in your own words, what does it mean for something to be inconsistent under this heuristic?
 - b) Will your primary research for this heuristic come before or after you run the tests?
 - c) What sources of information should you research and how will you find them?
 - d) What counts as an “inconsistency”?
 - e) Is this particular inconsistency bad, good or neutral?
 - f) Can we determine whether a particular inconsistency is minor or serious?
- 3) Now please give feedback to the other students who answered these questions.
- 4) If you have time, please look over the posts in at least one other discussion (analysis of one other heuristic) and give feedback / ask questions in that discussion as well.