Measuring Change and Coherence in Evaluating Potential Change in View

Gilbert Harman

Cognitive Science Laboratory
221 Nassau Street
Princeton University
Princeton, NJ 08542
(609) 987-2819
uucp: princeton!mind!ghh
arpanet: ghh@mind.princeton.edu

Marie A. Bienkowski

Bell Communications Research
Morristown, New Jersey

Ken Salem

Department of Computer Science
Princeton University

Ian Pratt

Department of Philosophy
Princeton University

ABSTRACT

In changing your view, you must balance the amount of change involved against the improvement in explanatory coherence resulting from the change. Even if change and improvement in coherence are measured by simply counting, there can be no general requirement that the number of modified items (added or subtracted) be no greater than the number of new explanatory and implications links. The relation between conservatism and coherence is more complex than that.

Keywords: explanation, planning, reasoning.

The purpose of this note is to discuss one aspect of the approach to the theory of reasoning described in Cullingford, Harman, Bienkowski, and Salem (1985), Harman (1973, 1986), and Harman, Cullingford, Bienkowski, Salem, and Pratt (1986). In this approach, reasoning is identified with change in view, that is, with additions and subtractions to your antecedent beliefs and plan. We suppose that among the most basic principles of change in view are (1) *conservatism*: other thing's being equal, minimize the amount of change; and (2) *coherence*: other thing's being equal, maximize the explanatory coherence in the resulting view. In order to apply these principles, you need a way of measuring how much change a suggested

The research reported here was supported in part by a research grant from the James S. McDonnell Foundation by the Defense Advanced Research Projects Agency of the Department of Defense and by the Office of Naval Research under Contracts Nos. N00014-85-C-0456 and N00014-85-K-0465; and by the National Science Foundation under Cooperative Agreement No. DCR-8420948 and under NSF grant number IST8503968. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the McDonnell Foundation, the Defense Advanced Research Projects Agency, or the U.S. Government.

modification would involve and how much increase in coherence. We are exploring very simple measures which simply count the number of changes and number of explanations a modification would involve. We considering exactly what it would be plausible to count and also what other principles are needed if we are to use such simple measures. In the present note, we will discuss only one aspect of the issue.

Other basic principles of reasoning include (3) *no get back*: avoid giving up things that you can infer right back; and (4) *clutter avoidance*: do not clutter your mind with trivialities (Harman 1986, chapters 2 and 6). We are not concerned with these principles in the present note. Rather, we want to consider what has to be true of reasoning if the impact of the amount of change and coherence involved can be measured simply by counting changes and explanations.

One possible outcome might be that there is no way to make this work. (The present authors disagree about whether this is the most likely outcome.) Perhaps we have to assign *weights* to propositions and explanations in such a way that some propositions and explanations count for more than others in calculating how much change in view or coherence would be involved in a particular proposed modification. It is even possible that there is no interesting way in which considerations of amount of change and coherence can play a useful role in deciding what revision to make in your plans. But let us assume as a working hypothesis that such considerations are relevant and can be measured by counting. The question then is what else might be involved in a system of revision that these considerations can be measured by counting.

Clutter avoidance requires that an acceptable change in view must promise to advance your goals. You start with an interest in whether such and such is true or an interest in having or doing something. Such interests can give you other interests, e.g., an interest in whether some other thing is true or an interest in having or doing something else. A new conclusion is acceptable only if its acceptance promises to satisfy one of your interests. For present purposes we may assume this means that a new conclusion is acceptable only if it contains a proposition P in which you are interested or want to be the case.

Comparing New Items with New Explanations

Consider an elementary logical inference. You start with a beliefs in "P" and "if P then Q" and infer "Q". We suppose that this is to infer an explanation: "Q because P and if P then Q." "Q" is the *explanandum* of this explanation and "P" and "if P then Q" are the *explainers*. (The sense in which an implication can be explanatory is discussed in Harman 1986, chapter 7.) How do we count the amount of change involved in this case and how do we count the added coherence?

We might say that this involves one new belief, namely "Q," and one new explanation, namely, "Q because P and if P then Q." In that case, the number of new beliefs is the same as the number of new explanations. The loss to conservatism involved in adding the new belief is matched by the gain in coherence involved in the new explanation. Clutter avoidance allows you to make an inference in this case only on the assumption that you are interested in whether Q.

This example may seem to suggest a requirement on acceptable changes that the number of things accepted must not be greater than the number of new explanations.

That would not mean you need an explanation of every new belief you accept. Often you accept a belief because it explains something else you already believe ("inference to the best explanation"). In that case, you do not need an explanation of the new belief. It is enough that in accepting that new belief you come to have a new explanation of something else.

The principle that the number of new things accepted must not exceed the number of new explanations would rule out arbitrarily inferring Q from any random belief P. Such an inference would involve the acceptance of one new belief with no new explanation. So, it would be ruled out by the principle as stated.

However, there is an objection to this way of counting and so to the principle that the number of new things accepted must not exceed the number of new explanations. In the first example, you come to accept not only "Q" but also that there is a certain connection between that belief and two other beliefs you already accept. That is, you come to accept that these two other beliefs imply "Q". There are really two new beliefs in this case and only one explanation. So it may seem that the principle that the number of new things accepted must not exceed the number of new explanations would prevent you from inferring "Q" from "P" and "if P then Q." In that case, the principle would be clearly unacceptable.

Now this implication is an instance of modus ponens. All such instances are immediately intelligible. In supposing that "Q" is implied by "P" and "if P then Q," you are not left wondering why there is this implication. You immediately understand the relation as implicational. You have all the explanation you need of the implication.

This is not to say that the implication is explained by the general logical principle of modus ponens. To say that would simply push the problem back one step (Carroll 1956). If we said that the implication holds because it is an instance of modus ponens, which you already accept, we would have to say that you accept three new things, (i) "Q," (ii) '"Q" is implied by "P" and "if P then Q",' and (iii) "the preceding belief (ii) is an instance of modus ponens." So, again, there would be more new beliefs than explanations unless one of the new beliefs needs no explanation. At some point you accept things that need no further explanation.

This suggests modifying the principle to say that the number of new things accepted *that require explanation* not be greater than the number of new explanations.

That would allow two methods of counting the number of changes. Either count the total number of new beliefs, including those that need no explanation, or count only the number of new beliefs that do not require explanation. Clutter avoidance favors counting all new beliefs, including those requiring no explanation. We do not want to accept even those beliefs without some special interest in whether they are true. Otherwise, there would be nothing wrong with cluttering your mind with trivialities of this sort.

This method of counting, along with our modified principle, would prevent the inference of "Q" from arbitrary "P" via the expedient of also inferring "P implies Q." If we were to count this as one new belief and one new explanation of that belief, using the initial method of counting, the inference would not conflict with our original principle. Contrary to the initial method of counting, it is obvious that two new things are accepted in this case. The inference conflicts with the modified principle, since the inference appeals to a connection between "P" and "Q" that requires an explanation. The connection needs an explanation since it is not immediately intelligible in the way in which instances of modus ponens are immediately intelligible. So the inference would involve the acceptance of two new beliefs requiring explanation and only one new explanation, which would violate the principle that the number of new things accepted that require explanation should not be greater than the number of new explanations.

It may seem that there is still a way to infer an arbitrary conclusion without violating this principle. From arbitary "P" you infer a complex e-structure containing two explanations. One explanation explains "Q" by appeal to the two explainers "P" and "P if and only if Q". The other explanation explains "P" by appeal to the two explainers "Q" and "P if and only if Q." Here there are only two new beliefs not requiring explanation, namely, "Q" and "P if and only if Q." The two implicational links are both immediately intelligible and therefore need no explanation. There are also two explanations in this case. Since the number of explanations is as great as the number of new beliefs needing explanation, this would not be ruled out by the principle that there should be a new explanation for every new item accepted that needs an explanation alone.

Notice that the complex e-structure you would accept in this case would be *circular*: Q would be taken to be explained in part by P, and P would be taken to be explained in part by Q. So, ultimately, Q would be used to explain itself. If that were allowed, you could even more simply infer an arbitrary "Q" by accepting "Q because Q." Here there would be only one new belief needing an explanation, namely, "Q" and there would be one "explanation," namely, "Q because Q." But that is no explanation; it is blatently circular.

In other words, the principle that the number of new things accepted that require an explanation should be not exceed the number of new explanations presupposes the following principle:

No circular explanations: acceptable e-structures must not contain circular explanations. If a proposition is explained in the e-structure, it must not itself serve as an explainer in the e-structure of one of its own explanatory antecedents. (It cannot be one of its own explainers, an explainer of one of its explainers, an explainer of an explainer of one of its explainers etc.)

The Flow of Acceptability from Prior Beliefs

But there is a more serious problem with our suggested principle that the number of new accepted items requiring explanation be no greater than the number of new explanations. This principle clearly fails in many cases of inference to the best explanation.

For example, Albert tells you Jack is a philosopher. You infer that Albert says this because he believes Jack is a philosopher and wants you to know whether Jack is a philosopher. Here, using the original method of counting new beliefs and ignoring new beliefs about explanatory connections, there are two new beliefs -- that Albert believes Jack is a philosopher and that Albert wants you to know whether Jack is a philosopher -- but there is only one new explanation, namely that those two things explain Albert's saying what he says. There are more new beliefs than explanations in this case, so the proposed requirement would say that this e-structure is not acceptable. But an inference to such an explanation is often quite in order even though it involves more new beliefs than explanations.

Someone might argue that the preceding example is acceptable only given something like the background belief that normally, if someone says something, then that is because the speaker believes what he or she is saying and wants the hearer to know. It might be argued that this background belief helps to provide further explanatory links between Albert's saying that Jack is a philosopher on the one hand, and, on the other hand, his belief that Jack is a philosopher and his desire that you should know that. We certainly must allow for explanations that explain why a given A is a B by noting that normally A's are B's (Harman 1986, Chapter 7). The suggestion is that you infer that a certain explanation holds on this occasion because normally, when someone says something, that sort of explanation holds. This analysis requires distinguishing two explanations. First, there is the psychological explanation of Albert's remark that Jack is a philosopher, an explanation which appeals to an assumed belief and desire of Albert's. Second, there is something like a statistical explanation of that psychological explanation's holding in this case, an explanation that appeals to what normally leads to someone saying what he or she says.

It may look as if this analysis of the inference would violate the prohibition against circular explanations, but it does not really do so. The appearance of circularity arises because you end up explaining why Albert believes that Jack is a philosopher in part by appeal to Albert's saying that Jack is a philosopher and you end up explaining Albert's saying that Jack is a philosopher in part by appeal to Albert's believing that Jack is a philosopher. But in accepting this inference you would not have to give any credit to this last explanation. The coherence in your views that it contributes would not have to counted in order to determine that your inference is acceptable. This follows from the fact that this analysis takes each one of Albert's new beliefs to be explained, so the number of new beliefs cannot exceed the number of explanations. Albert's remark and the principle about what normally explains such remarks together are taken to explain: "Albert says Jack is a philosopher because Albert believes this and wants me to know it." And the truth of the proposition just quoted obviously implies both "Albert believes that Jack is a philosopher" and "Albert wants me to know that Jack is a philosopher." So, there are at least as many explanatory and implicational connections as new beliefs requiring such connections in this analysis without counting the connection provided by explaining what Albert says via his belief and desire.

But there are problems with the analysis. It analyzes away the inference *to* an explanation, turning it into an inference *from* an explanation. Inference to an explanation seems to be a genuinely different sort of case. In many cases of inference to the best explanation there are no relevant prior beliefs about what normally or probably happens. Furthermore, if the suggested analysis were correct, there would be no real need in the present case to infer an explanation of what is said. You could, for example, infer from "Albert says that Jack is a philosopher" to "Jack is a philosopher" via the default principle, "Normally, when someone says something it is true," without having a view about how Jack's being a philosopher might be part of the explanation of Albert's saying this. This would make it difficult to account for the way in which such an inference is defeated by the discovery that Albert is insincere or that he is not in a position to know whether Jack is a philosopher (Harman, Cullingford, Bienkowski, Salem, and Pratt 1986).

If the analysis via a default rule does not work for the general case, as we are suggesting may be true, then we have to abandon the rule that there should be at least as many new explanations as new things accepted that require explanation. What might replace that principle?

Here's a proposal. Consider the case in which the relevant change in view being considered is entirely a matter of adding new things. The things added must be represented as a non-circular "e-structure," that is, as a set of "e-nodes" or simple explanations, each of which has one or more "explainers" and a single "explanandum" (the thing explained or implied). Some e-nodes themselves may be explananda in other e-nodes (for example, if a given explanatory connection is intelligible as an instance of a principle about what might explain what). We can say that an explainer or explanandum in one of these e-nodes or the e-node itself is "OK" if (but not only if) it is either already accepted or immediately intelligible (or obvious). Furthermore, if the explanandum of an e-node is OK, then all the explainers are OK; and, if all the explainers in an e-node are OK, so is the explanandum. Finally, the whole e-structure is "acceptable" only if all of its contained e-nodes and their explainers and explananda are OK. (Clearly, it is trivial to check whether this condition is satisfied.)

When you infer from "P" and "If P, Q" to "Q," the relevant e-structure has a single e-node, whose explainers are "P" and "If P, Q" and whose explanandum is "Q". Since both explainers are previously accepted, they are OK; so the explanandum is also OK. The e-node is OK since it is immediately intelligible as an instance of modus ponens. All the items in the e-structure are OK, so the e-structure as a whole is acceptable.

The simple inference from arbitrary "P" to arbitrary "Q" would be ruled out because it involves no e-nodes and so allows no way for "Q" to count as OK. This inference could not be made acceptable by adding an arbitrary e-node connecting "P" as explainer with "Q" as explanandum, since this e-node would not be OK. Nor would it help to add two e-nodes, one with explainers "P" and "P if and only if Q" with explanandum "Q" and the other with explainers "Q" and "P if and only if Q" with explanandum "P", since the resulting e-structure is circular.

Our example of inference to the best explanation is now clearly acceptable. When you infer that Albert says Jack is a philosopher because he believes it and wants you to know, the main e-node has as its explainers "Albert believes Jack is a philosopher" and "Albert wants you to kow whether Jack is a philosopher." Its explanandum is "Albert says that Jack is a philosopher." Since you already accept the explanandum, it is OK. Since the explanandum of that e-node is OK, the explainers are also OK. Finally, the e-node itself is OK, since it is intelligible as an instance of a default principle which you accept, namely, "The belief that something is so plus the desire to tell someone whether it is so can lead one to say that it is so."

When several acceptable e-structures compete (i.e. have conflicting elements), a particular e-structure can be inferred only if it is the best of the competing acceptable e-structures. We are left with the need for a way of evaluating the e-structures e.g. by counting the amount of change each involves and comparing it with the amount of coherence it brings to your overall view. We must leave for further discussion whether we should count the total number of new things accepted, the total number of new e-nodes accepted, or something else.

In any event, this last approach does not take conservatism to be just a matter of minimizing the total number of new beliefs. You also need to consider which beliefs in your projected modified view were previously accepted. In assessing an e-structure, you also need to note which of its elements are things you already accept and you need to consider whether the acceptability of those items flows over to all the elements of that e-structure in accordance with the principle just given.

Simple Plans

We conclude by considering a very simple case of practical reasoning. You want to raise your arm, where that is something that is immediately within your power. So, you decide to raise your arm. We are supposing that this involves the acceptance of the following explanation: "I will raise my arm because of my decision to raise my arm."

Here there are the following new beliefs: (1) "I will raise my arm," (2) "I decide to raise my arm," (3) (1) because of (2). (1) requires an explanation and an explanation of (1) is accepted, namely, (3). Let us suppose for the moment that we do not have to worry about (3). Either we can suppose that (3) does not require an explanation, because raising your arm is something you take to be immediately within your power. Or, we can suppose that (3) requires an explanation and is explained by something already

accepted, namely that raising your arm is immediately within your power. We will come back to this below.

What about (2), "I decide to raise my arm"? Since you have a reason for deciding to raise your arm, we could suppose that your plan involves the acceptance of the idea that you decide for that reason. For example, your plan includes the thought that, because of your desire to raise your arm, you accept this very plan, which involves deciding to raise your arm, which leads to your raising your arm.

Your plan has to include a reference to the reasons for it in order to allow you to allow the plan to be abandoned if the reasons for the plan are no longer applicable. Consider a case in which you adopt a complex plan in order to obtain a goal G. You plan to do M1, which will put you in a position to do M2, which will put you in a position to do M3, and so forth, so that you are in a position to do M11, which will get you G. While in the midst of carrying out this plan, as you are doing M3, you learn that doing M11 will not get you G after all. At this point, you want to be able to abandon the whole plan so that you no longer intend to do M4, M5, and so forth. Just how you are able to do this is something we must eventually consider, but it is clear that you will be able to abandon these intermediate actions only if you keep a record as to why you are undertaking them.

Another reason to record that a plan is aimed at satisfying a particular desire, is to have a way to prevent that desire from leading to the development of other plans designed to satisfy that desire. Once you have a plan to attain a certain goal, you do not have to look for another way to attain that goal!

So, it seems that in a very simple case of deciding to raise your arm you accept a rather complex plan: "In order to satisfy my desire to raise my arm, I am led to adopt this plan, which involves my deciding to raise my arm, which leads to my raising my arm." Consider the various things this involves. First, there is the information that you desire to raise your arm. That (we may suppose) is something you already accept. (We discuss below how your acceptance of that belief might satisfy the principles of change in view.) Second, there is your recognition that you adopt the whole plan. You take your adoption of the whole plan to be explained by your desire to raise your arm. Third, there is this assignment of an explanatory link, the thought that your desire leads you to adopt the plan. We can suppose that that connection is immediately intelligible and therefore needs no further explanation. Fourth, there is the thought that you decide to raise your arm. You take that to be explained by your adopting the plan. Fifth, there is this last explanatory connection between your adopting this whole plan and the decision to raise your arm. The existence of that connection needs no explanation, since the adoption of that decision is obviously part of the plan. Sixth, there is the thought that you do raise your arm. You take your raising your arm to be explained by your decision to raise it. Finally, seventh, there is the explanatory connection between your decision to raise your arm and the fact that you raise your arm. This needs no explanation, because it is obvious to you how the decision leads to the raising.

Summary

In changing your beliefs and your plans, you accept not only simple beliefs and plans but also implications among these simple beliefs and plans, explanations of them, as well as implications among and explanations of implications and explanations. In considering whether a change in view is even minimally acceptable, it is necessary to keep track of which of its elements are already accepted and whether acceptability can flow from these elements to the whole of the proposed new e-structure. It may be that competing e-structures can be judged in part on the basis of counting the changes they involve, but there is no general requirement that the number of new items needing no explanation must be no greater than the number of new explanations.

On another occasion we will extend this analysis to include cases in which change in view involves giving up something previously accepted.

Bibliography

Carroll, L, (1956) What the Tortoise Said to Achilles. *The World of Mathematics, Volume 4*, edited by Newman, J. R. New York, Simon and Schuster. Pp. 2402-2405.

Cullingford, R. E., Harman, G., Bienkowski, M, and Salem, K. (1985). Without Logic or Justification: Realistic Belief Revision, *Proceedings of the National Academy of Sciences Workshop on Artificial*

Intelligence and Distributed Problem Solving. Washington, D. C.: National Academy of Sciences.

Harman, G. (1973) *Thought* Princeton, New Jersey. Princeton University Press.

Harman, G. (1986). *Change In View: Principles of Reasoning*. Cambridge, Massachusetts. M.I.T. Press.

Harman, G., Cullingford, R. E., Bienkowski, M. A., Salem, K., and Pratt, I. (1986) Default Defeaters in Explanation-Based Reasoning, *The Eighth Annual Conference of the Cognitive Science Society* Amherst, Masachusetts, Lawrence Erlbaum, pp. 283-291.