PHIL 323 Keith Douglas
Prof. Andrew Irvine 13714993

Gärdenfors' Belief Dynamics

Introduction:

Peter Gärdenfors' (1988) book *Knowledge in Flux: Modeling the Dynamics of Epistemic States* focuses on modeling the belief changes of an idealized rational agent. In chapter three of this work he introduces a set of axioms purporting to describe the expansions, contractions and revisions of beliefs of this agent. The present paper will first give a brief overview for the axioms presented for expansions and some more general postulates used throughout his text. It will next discuss each of these in turn, discussing each's plausibility in light of formal and factual considerations. Particularly close attention will be paid to a notion of minimality which Gärdenfors develops; this notion is sufficiently problematic to merit a section on its own.

Section One: Expansion Axioms Overview

In order to understand the axioms for expansion, one first must look at the background formal machinery which is being used by Gärdenfors (1988).

Chapter two introduces the basis of the model he will be using, namely the belief set and its contents. A belief set is simply a set of propositions assumed to be formalized in a fashion much like a standard propositional logic, and gives an overview of symbols he uses, which I have tried to faithfully reproduce as much as possible. He introduces the symbols $\top$ and $\bot$ to represent "truth" and "falsity"[1] mainly to discuss what he calls the absurd belief set. This is a[2] belief set containing a contradiction, and hence, as we shall see, 'everything possible'. Beliefs are represented in this logical

[1] Gärdenfors also claims (pg. 21-22 1988) that these are not truth values. It is not clear to me what "truth" and "falsity" are if not truth values. Even an extreme relativist will accept that they are; she will just simply deny all of the traditional ways they get assigned.

[2] Actually, as we shall soon see, all absurd belief sets are identical on the assumption of logical closure of belief sets.

language, and collections of them form belief sets. He then introduces the three attitudes he is considering that an agent may have towards any particular "belief-sentence" A. A can be in the particular agent's belief set; he calls this acceptance. A may also be rejected by the agent, which is to say that the agent believes -A[3]. Finally, A may neither accepted nor rejected by the agent: A is indetermined.

Since Gärdenfors is dealing with rational[4] agents alone, there are three postulates to govern this assumption of rationality. These are that the sentences accepted by an agent are to be logically consistent, and that logical consequences of what is accepted should also be accepted. This has the side effect of creating (denumerably) infinite belief sets once an agent has a non-empty belief set[5]. (I stress this as it will be important later.)

Gärdenfors calls these the postulates of informational economy and the two postulates of rationality. The informational economy postulate can be stated as: when changing the contents of belief sets one ought to change as little as possible. This higher-order postulate is supposed to motivate several of the more substantive axioms (in particular numbers three and five of the set of expansion axioms). The two postulates of rationality are that as far as possible, belief set consistency is desirable, and that all logical consequences of beliefs in a given belief set are also in that set.

Gärdenfors also stresses that each belief is accepted to the same degree of

[3] I note in passing that this may prove problematic if one is an intuitionist in logic - Gärdenfors assumes throughout that ∂--A ↔ A.

[4] That is, logically rational agents alone - Gärdenfors does not draw any distinctions between possible meanings of rationality. It could be argued that Gärdenfors is acting prematurely by starting modelling with a logically rational agent and not a semantically rational agent. See Bunge 1999, entry "rationality", for the ordering of eleven distinct notions of rationality.

[5] Proof: Assume a belief set K={a}. This belief set, by the logic Gärdenfors allows to govern the belief states (pg. 24) permits the following derivation: a∂a. Hence, the agent can derive by the deduction theorem ∂aÇa and so on to aÇaÇa, etc, each a logical consequence of the single belief hence also in the new belief set, *ad infinitum*. For a slight possibility that there could be a nondenumerably infinite belief set, see footnote 10.

credence ("probability"[6]). The remaining sections of chapter two concern other possible models of belief states; they need not concern us here.

In chapter three the axioms for belief changes are presented within the framework of belief sets developed in chapter two. Expansion is represented as a function from the cartesian product of the belief set and the set of setences. Six postulates govern the behaviour of this function, symbolized "+". These are as follows, using his symbolism of letting K_A^+ represent "belief set K is expanded by belief sentence A". Let H also be a belief set.

1. K_A^+ is a belief set. In other words, belief sets are closed under expansion.

2. $A \in K_A^+$. I.e. the input "A" to a belief set under expansion can simply be viewed as the sentence which "produces" the expansion.

3. $K \subseteq K_A^+$. The set of all beliefs prior to an expansion is a subset of the new belief set. In other words, expansion does not cause an agent to "lose" any beliefs.

4. $A \in K \rightarrow K = K_A^+$. An existing belief within a belief set K being added to K does not change K.

5. $K \subseteq H \rightarrow K_A^+ \subseteq H_A^+$. This postulate insures that if two agents with belief sets H and K are such that one believes everything that the other did and possibly then some, this relation still holds after each expands by A.

The final postulate for expansions alone is:

6. For all belief sets K and all sentences A, K_A^+ is the smallest belief set that satisfies 1-5 above.

[6] "Subjective probabilities" have their share of problems (see, e.g., Bunge 1996, 1999), and Gärdenfors would be correct to stay clear of this "hornet's nest". It is not problematic to survey the literature on the subject, as he does, of course.

Now that we have seen the six expansion postulates and three more general ones, we shall now turn to a discussion of these. Postulates one through five will be taken up here; postulate six will be taken up in the third section of the present paper.

Section Two: Discussion of Postulates

This section consists of a brief discussion of each of the first five of Gärdenfors' postulates, particularly in light of the three general postulates.

Postulate one seems to be relatively straight forward and require little commentary. Postulate two and three are a bit problematic. Gärdenfors' model assumes that all beliefs are equally accessible or subjection to revision[7] to the agent; there are not even any "hinge propositions" à la Wittgenstein (1969) or "temporary foundations" à la Bunge (1983)[8]. I regard this as too much of an idealization for the model to have much use. A purely logical agent, even one who does have access to all the logical consequence of her own beliefs (as Gärdenfors' second rationality postulate proposes), cannot even act, as the decisions one makes are not normally logical, though of course they (usually?) presuppose logic. Note that this point holds no matter how "bare-boned" one's model of belief dynamics is supposed to be, as it makes the whole enterprise of dynamics somewhat empty.

The gist of this is that logic alone can never be the basis for what specific

[7] One need not not hold that one's "most important" beliefs are least subject to revision, as is commonly held, in order to have "revisability" rankings. Conceivably one could have beliefs that are harder to revise because they are "harder to bring to memory" or the like. (This could even include "require more blood to the brain than is available right now" - conceive of a drunk - who cannot change her beliefs as easily as someone sober - for something like that reason.)

[8] I use these two examples as they are what one might call "foundations without foundationalism". Bunge's thesis is that certain postulates are transient and revisable but are still pragmatically, morally, and especially epistemologically justificatory of other beliefs. He feels that this captures the model of science well; hence the view is part of his account of scientific realism. Wittgenstein's hinge propositions are a sort of "ordinary language" version of this same thesis, at least on one reading.

beliefs to reject; that's why the "never losing any beliefs by expansion" maxim and its associated postulates is unrealistic. Consider the case of someone who has two basic beliefs[9]. Put K={A→B,-B} as this set. Now, the agent encounters "A". Should she believe it? This is not a matter of logic, in general. (It would be if the belief under consideration were about logic, which presents an interesting self-reference problem[10] of sorts that I need not go into.) We expand her belief set to {A→B, -B, A} then draw the logical consequences, producing a pair of inconsistent beliefs {B, -B}. She can reject[11] A, then, reject -B, or she can reject A→B. There's no logical reason for any of these choices. Hence saying that 'expansion' does not require deletion of anything does not work because it may require some in order to maintain consistency[12]. This requires updating all logical consequences in such a way as that this "no deletions" axiom is violated. As we have seen, Gärdenfors has asserted that the agents modeled by his system believe all the logical consequences of their beliefs, thus producing an (countably) infinite set of beliefs. In fact, changes caused by an update of the sort discussed above, then, would cause an infinite number of deletions, the apparent

[9] I am using "basic beliefs" as the set of beliefs of an agent which is a minimal set of beliefs necessary to derive all and only the other beliefs of the agent logically. (Note in particular that in the case of an agent with the absurd belief set, a single contradiction suffices.) As I am going to take issue with Gärdenfors' use of minimal later, this use should be kept in mind, particularly the fact that it is not unique. Proof: Put K = {a→b}. Then by contraposition one can expand to get K={a→b;-b→-a} which allows us to make another minimal set K={-b→-a}. There is a problem with this notion which I will discuss later, but it has intuitive appeal, so I make use of it here.

[10] In particular, the problem would arise if the logic that governed the structure of the belief sets was itself a subset of beliefs contained in the belief set, i.e. if the agent had access to her own (logical) belief mechanisms as part of her beliefs. (Fo example, the agent believes that her beliefs are closed under *modus ponens*, and they are or are not so closed.)

[11] Here "reject a belief α" means "convert α to undetermined" or "accept $-\alpha$". Nothing hinges on the choice.

[12] It may be argued that this would require an iterated update of sorts - that after every expansion a revision process has to occur. This possibility is acceptable, but it reduces some of the axioms for expansion to be somewhat pointless if this revision process were to occur "immediately", as it must to avoid having the agent "pollute" her belief set by immediately drawing all the consequences from a contradiction produced.

inconsistency or “weak point” in Gärdenfors’ account noted above on a grand scale.

This is where the postulate of informational economy discussed above seems to reflect something of my first intuitions on the matter. However, the idea that one should change the bare minimum necessary is of course trivially fulfilled in this case, as one can simply reject what was recently added to one’s belief set. In the above example, this suggests our agent should just drop “A”. However, this does appear unsatisfactory for several reasons, not the least of which is the epistemic conservatism involved in always dropping the potentially new belief addition to save consistency at all times. It appears to me at this stage that there is no general way or principle by which one could decide what to accept and what to reject in each case without having credences associated with at least some of the beliefs in the belief set. (Any principle that said “believe what is necessary and no more!” would be an uninformative way of stating such a principle in the absence of an explanation of “necessary”.) Note also the problem with dropping in general. If a belief B is in not some belief set K, an agent undergoing K_B^+ will in general add an infinite number of new beliefs along with this one, so in what sense “minimality” is to play a role here is unclear. See section three below for an exploration of the minimality issue in another context.

Postulate four is fine (granting the assumption that beliefs are binary and have no credences associated with them, of course). Postulate five is worrisome for the same reason that postulate three was. If we have two agents adding to their belief sets, both of which are consistent to begin with, but both introduce a belief that allows an inconsistent set to be generated[13], no logical decision can be made on what to subsequently delete to restore consistency. Hence it is *prima facie* possible that one agent will delete more beliefs to regain consistency than and another, and thus delete enough so that the belief set containment relation that was held previously is reversed[14].

[13] In other words, the raw belief plus the new logical closure “cause” the inconsistency. This is different from an agent who’s belief set is (e.g.) K = {A} then expands by -A.

[14] Example: Put $K_1 = K_2 = \{a \rightarrow b;\ -b\}$ as one way of expressing the basic beliefs

This brings us to the worries about the minimalism of postulate six.

Section Three: Problems with Postulate Six

Postulates three and five are not the only place where the infinite nature of belief sets gets Gärdenfors in trouble. There is in addition a conflict with postulate six and the logical closure axiom. Since all belief sets are infinitely large, it is not immediately obvious in what sense "minimality" is to be ascribed to belief sets. Gärdenfors himself (section 3.5, pg. 66) that some sort of information-theoretic notion could be used. We shall briefly see that this off-handed suggestion is at best incomplete. Hence we shall look at two other ways in which to understand this postulate. Cardinality is one way, but as previously stated, the cardinality of all belief sets is $\aleph_0$. This entails that a direct comparison of size is not apt. Nevertheless there is a possible avenue to explore here which may have occurred to some which expands on the notion of a basic belief set explored in section two of the present paper, but is plagued with problems of a different character. We shall also examine a "subset comparison" method which seems to work and capture some of Gärdenfors' intended views, but is rater ad-hoc.

Section 3.5 of Gärdenfors 1988 discusses briefly Gärdenfors' views on his term "minimal change". He writes that information-theoretic notions can be used to discuss expansions and deletions from belief sets in a "probabilistic" context. There are two problems with this view. One is that he has not as yet made his model "probabilistic"; the second concerns how exactly one could apply information theory (IT); without details (and none are found in chapter 5, either, where he discusses IT again briefly.) this is just a promise for future work. (A skeptic might point out that the notion of using subjective probabilities in place of objective ones in IT appears somewhat dubious[15].)

(this term is defined in footnote 5 above) of two agents. Clearly $K_1 \subseteq K_2$ as required for the fifth expansion postulate. Then expand each belief set by a; it follows that each agent may revise differently. Suppose $K_{1\ r}$ rejects the new a after all; so still $K_1 = \{a \rightarrow b;\ -b\}$. But K_2 gets rid of the conditional and now believes {a, -b}. But now, after the expansion (precisely because expansions may require other modifications to belief sets in order to save consistency), $K_1 \not\subseteq K_2$.

[15] Subjective probabilities used in an IT context would make the information

His other proposal, for non-"probabilistic" contexts involves comparisons of size. We shall see, below, that this does not work. The present author has nothing further to say on these proposals as there is nothing further substantive on them available. I will now turn to two possibilities Gärdenfors does not consider.

As we have seen, if one introduces the notion of basic belief set (footnote 9 above), one can try to use this notion to compare belief set size. This requires an additional general postulate for the whole system (not merely to expansions) of belief dynamics Gärdenfors has developed. This new postulate would be that the basic belief sets under consideration would be finite. This latter restriction is not quite as innocuous as it sounds. Often times one might associate infinite sets of basic beliefs with a god or (perhaps) Laplace's demon and hence pointless to consider in a discussion of human belief dynamics. But depending on the form of beliefs within the set, it is not too inconceivable that some lowly humans have an infinite number of beliefs in some sense as well. Consider a human alone, out in the intergalactic void. She "looks out into infinity" past the end of her arm, and sees "how far out it goes". Does she now believe that "nothing at location x, x+h ..."[16]. This does not seem likely, but it does pose the question of how beliefs are to be represented to avoid these sorts of worries. One could argue that the agent need not believe each of those sentences; one would simply have to believe "nothing from here on", or at least "nothing in some spatial interval [x, x+∂x]." If the latter is a continuous interval (see note 15 below) then even the latter involves an infinite number of potential content of the system concerned (here, the belief set) subjective as well, which seems dubious.

[16] First, I selected the intergalactic void example to show that there need not be much of anything at all (just her body!) in the world around an agent to bring this intuition to light.

Second, if one does not think spacetime is discrete, as I have implied, this creates an even graver problem, and one that I think is rather absurd - the possibility of someone with a nondenumerable infinity of beliefs. (Probably Laplace's demon in a continuous world would have such, as it may very well be able to predict the exact gravitational/magnetic/electric field produced by something at any time, and certainly omniscient gods would, as well.) Gärdenfors could presumably try and overcome this problem by asserting that because he has conceived of each belief as a sentence there can only be 'at most' a countable infinity of them.

beliefs[17].

Assuming, then, that one can avoid infinite basic belief sets by whatever means, then there is in principle appears to be a way to discuss minimality using the notion of the cardinality of set after all. This method would work as follows in the context of axiom 6 above: if two possibilities exist for expanding an agent's belief set, and one (say, the first) involves a smaller basic belief set than the other after the expansion, then the principle of minimal change suggests the first set be the new belief set.

However, there is a grave problem with this at first glance very plausible suggestion. Since Gärdenfors has (1988, pg. 25) defined a generalized conjunction over infinite sets of propositions, all basic belief sets can be reduced to a singleton[18], assuming they are not nondenumerably infinite.

This suggests that the notion of minimality as understood in a basic belief set sense is rather useless. So I will now turn to a set-inclusion notion of minimality.

This idea is as follows. A belief set K with beliefs $\{b_1 \ldots b_n\}$ is the minimal belief set of beliefs b_1 to b_n if for all belief sets H including $b_1 \ldots b_n$, $K \subseteq H$. Then uniqueness of this set is a problem. One example noted previously is that $\{a \rightarrow b\} \neq \{\text{-}b \rightarrow \text{-}a\}$ and yet in some sense the two belief sets are equivalent. One could also have the belief set {-(a & -b)} and any number of others which even use different connectives[19]. This suggests the

[17] The way of stating the latter brings out yet another worry that one might have about Gärdenfors' project more generally that I have no time to develop in the present paper. This concerns more semantic inferences between beliefs, and where one draws the line between semantic inference and purely logical inference. Gärdenfors has stated that the belief sets he is considering are governed only by first order logic, but suppose an agent believes (*inter alia*) 2+2=4 and (say) Peano arithmetic. Do they then believe that 3+2=5? 4+2=6?

[18] Proof: start with some allegedly minimal belief set $K = \{a_1, a_2, \ldots, a_n\}$. Then by conjunction one can combine each belief pairwise until one has $K' = \{a_1 \,\&\, a_2 \,\&\, \ldots \,\&\, a_n\}$.

[19] I note in passing that psychological data (see Medin and Ross 1997) suggests that we do not "carry" our beliefs in propositional form because we are notoriously bad at interconverting amongst equivalent logical forms.

following *ad hoc* solution to guarantee uniqueness of minimal belief sets in the set-inclusion sense. First restrict the connectives in which beliefs can be stated, e.g. to $\rightarrow$ and -. Then one also needs to make sure that no redundancy is present; certain long belief sentences (e.g.: $(p \rightarrow (p \rightarrow p))$) are tautologically equivalent to shorter ones; reduce all to their shortest form keeping in mind the minimal connectives. Perform this for all possible belief sets under consideration; each one will then be simplified and minimal in a sense that is something like what Gärdenfors seems to have in mind. Hence when comparing minimality of belief sets one should perform this operation, then check cardinalities of the resulting belief sets. If the belief sets are of the same size, they are equivalent, and we have thus saved Gärdenfors' minimality at the cost of its purported uniqueness. A method that overcomes the *ad hoc* nature of this proposal and retains its "spirit" is not currently known to the present author.

Conclusion

We have noted several interrelated possible confusions and shortcomings in Gärdenfors' account of expansions and several possible ways to clean up some of them. Given the general nature of the problems, it does seem likely that similar difficulties plague the other axioms he presents. Hence it seems somewhat implausible that he has successfully captured even a basic account of belief dynamics.

Bibliography

Bunge, M. 1983. *Treatise on Basic Philosophy, Vol. 6: Understanding the World*. Dordrecht: Reidel.

Bunge, M. 1996. *Finding Philosophy in Social Science*. New Haven: Yale University Press.

Bunge, M. 1999. *Dictionary of Philosophy.* Amherst: Prometheus Books.

Gärdenfors, P. 1988. *Knowledge in Flux: Modelling the Dynamics of Epistemic States*. Cambridge: MIT Press.

Medin, D. & Ross, B. 1997. *Cognitive Psychology* (3e). Orlando: Harcout Brice College Press.

Wittegenstein, L. 1969. *On Certainty*. New York: Harper and Row.