Chapter 4

FALLACIES AS COGNITIVE VIRTUES

Dov M. Gabbay[1] and John Woods[2]

[1,2] *Department of Computer Science, King's College London*
dg@dcs.kcl.ac.uk

[2] *Department of Philosophy, University of British Columbia*
woodsj@dcs.kcl.ac.uk
jhwoods@interchange.ubc.ca

Sometimes you had to say Stuff Logic *and go with the flow.*
—Reginald Hill, *Good Morning Midnight*

Abstract In its recent attention to reasoning that is agent-based and target-driven, logic has re-taken the practical turn and recovered something of its historic mission. In so doing, it has taken on in a quite general way a game-theoretic character, precisely as it was with the theory of syllogistic refutation in the *Topics* and *On Sophistical Refutations*, where Aristotle develops winning strategies for disputations. The approach that the present authors take toward the logic of practical reasoning is one in which cognitive agency is inherently strategic in its orientation. In particular, as is typically the case, individual agents set cognitive targets for themselves opportunistically, that is, in such ways that the attainment of those targets can be met with resources currently or forseeably at their disposal. This not to say that human reasoning is so game-like as to be utterly tendentious. But it does make the point that the human player of the cognitive game has no general stake in accepting undertakings that he has no chance of making good on.

Throughout its long history, the traditional fallacies have been characterized as mistakes that are attractive, universal and incorrigible. In the present essay, we want to begin developing an alternative understanding of the fallacies. We will suggest that, when they are actually employed by beings like us, they are defensible strategies in game-theoretically describable pursuit of cognitive (and other) ends.

4.1 Introductory remarks

In its recent return to reasoning that is agent-based and target-driven, logic has recovered something of its historic mission. In so doing, it has taken on in a quite general way a game-theoretic character, precisely as it was with

O. Majer et al. (eds.), *Games: Unifying Logic, Language, and Philosophy,*
Logic, Epistemology, and the Unity of Science 15,

Aristotle's theory of syllogistic refutation in the *Topics* and *On Sophistical Refutations*. Aristotle here presents winning strategies for disputations. They pivot on the refuter's exploitation of his opponent's concessions. While the opponent must believe his concessions, the refuter need not. The approach that the present authors take toward the logic of agent-based target-driven reasoning is one in which cognitive agency is inherently strategic in its orientation. In particular, as is typically the case, individual agents set cognitive targets for themselves opportunistically, that is, in such ways that the attainment of those targets can be met with resources currently or forseeably at their disposal. This not to say that human reasoning is so game-like as to be utterly tendentious. But it does make the point that the human player of the cognitive game has no general stake in accepting undertakings that he has no chance of making good on. Throughout its long history, the traditional fallacies have been characterized as mistakes that are attractive, universal and incorrigible. In the present essay, we want to begin developing an alternative understanding of the fallacies. We will suggest that, when they are actually employed by beings like us, they are defensible strategies in game-theoretically describable pursuit of cognitive (and other) ends. Needless to say, the generically game-theoretic approach has developed several more specialized tendrils. Some of these involve a re-writing of classical first order logic. Others are extensions or adaptations of the mathematical theory of games. Still others refine the generic notion into technically versatile models of dialogue. All of these are welcome developments, and many are of enduring importance. In some of our writings in progress, the more peculiarly game-theoretic aspects of practical reasoning are developed. But we continue to think that the generic notion, embodying the fundamental idea of strategies for the attainment of cognitive targets, is also of lasting importance. This is something that we shall attempt to demonstrate in this essay.

The present work is adapted from our book in progress, *Seductions and Shortcuts: Fallacies in the Cognitive Economy* (Gabbay and Woods, 2009). Our principal purpose here is to introduce readers to that work's founding assumption, and to identify some of the considerations that lend the idea support. We also have it in mind to attend to an important ancillary matter. It is the task of elucidating the role of what an agent is *capable of* in assessing whether his performance is faulty or defective. The essay is structured as follows. In Part I we discuss the question of cognitive agency. Part II illustrates our approach to fallacies.

PART I: PRACTICAL AGENCY

We begin with the so-called Gang of Eighteen, the name given to a loose confederacy of presumed errors that are discussed with a considerable regularity

in the contemporary literature on fallacies (Woods, 2004).[1] In one recent treatment (Woods et al., 2004), the Gang of Eighteen is represented by the following list.

ad baculum
ad hominem
ad misericordiam
ad populum
ad verecundiam
affirming the consequent
amphiboly
begging the question
biased statistics
complex question
composition and division
denying the antecedent
equivocation
faulty analogy
gambler's
hasty generalization
ignoratio elenchi
secundum quid

The Gang of Eighteen (*GOE,* for short) embeds a certain view of what it is to be a fallacy. It sees fallacies as mistakes of reasoning (or arguing) that are attractive, universal and incorrigible. So conceived of, fallacies retain a striking kinship with Aristotle's original definition, in which a fallacy is an argument (or a piece of reasoning) that appears to be good in a certain way, but is not in fact good in that way. It is easy to see that the first two marks of fallaciousness are expressly caught by Aristotle's definition. For a fallacy is not only an error but, because it appears not to *be* an error, is a mistake that has a certain attractiveness. It is also clear that Aristotle intends the attractiveness of fallacies to give them a kind of universal appeal: Fallacies are errors that people in general are disposed to make, not just the logically challenged or the haplessly inattentive. If their attractiveness grounds their general appeal, it also grounds their incorrigibility. To say that a fallacy is an incorrigible error is to say that, even when properly diagnosed, there is a general tendency to recidivize. The modern notion incorporates these interdependencies. Accordingly, we have it that

[1] We emphasize the looseness of the grouping. In Copi (1986) 17 fallacies are discussed; in Carney-Scheer (1980) the number is 18; Schipper-Schuh (1959) runs to 28; and Black (1946) limits itself to only 11. While all these lists are pairwise inequivalent, there is nonetheless a considerable overlap among them.

Proposition 1 (Fallacies)**.** *A fallacy is a generally attractive and comparatively incorrigible error of reasoning (or argument).*

The *negative thesis* we wish to propose is that the general idea of fallacy is correct but that there is something gravely defective about the Gang of Eighteen and any of its standard variations. As we shall attempt to show, there are two difficulties with these lists:

1. Some of their members aren't fallacies.
2. Those that are errors aren't usually mistakes committed by beings like us.[2]

Our *positive thesis* is that

3. Several of the *GOE* are actually cognitive virtues.

To make good on these theses requires that

(a) We identify the members of the *GOE* of which the theses are true.

(b) Establish in each case that the relevant thesis is indeed true.

(c) Give some account of how it came to be the case that by our lights, the defective inventory of fallacies took hold.

In proceeding with these tasks we want to make it clear at the beginning that it is not our view that people don't commit fallacies. Our view rather is that the *GOE* has not managed to capture any of them in wholly convincing ways. For either they are indeed fallacies which we happen not to commit, or we do commit them, but they are not fallacies.

In its most usual meaning, a fallacy is a *common misconception.* It is an attractive, widely held belief that happens to be untrue. In many cases, it is also a belief that people have difficulty letting go of, even, after its falsity has been acknowledged. So whereas the received idea among logicians has been that a fallacy is an *argument* that is defective in the traditionally recognized ways, the view of the layman is that it is a *belief* that has the requisitely counterpart features. We may wish to take note of the point that if our present theses about the Gang of Eighteen can be sustained, we will have shown that the logician's inventory of the fallacies is in the layman's sense itself a fallacy.[3] If this should

[2]Given one's tendency to apply the word "incorrigible" to practices (or practitioners) one disapproves of, this is very much the right word for the fallacies *as traditionally conceived of.* Since ours is a view of the fallacies that rejects the traditional conception, we shall replace "incorrigible" with the more neutral-sounding "irreversible".

[3]A theme sounded by two recent writers. See Grootendorst (1987), which is entitled "Some fallacies about fallacies", and Hintikka (1987), which is entitled "The fallacy of fallacies". For reservations see, in the first instance, Woods (2004, Chapter 9) in the second, Woods and Hansen (1997), and for a rejoinder (Hintikka, 1997).

prove to be the right sort of criticism to press against the *GOE* approach, then something like the following argument schema must itself be defective. Let us call it

The Fallacy of Fallacies Schema

1. Practice *P* is universal, attractive and incorrigible (irreversible).
2. Practice *P* lacks property *Q* (e.g. validity).
3. Therefore, practice *P* is a fallacy.

Our view is as follows. There are members of the Gang of Eighteen of which (1) and (2) are true, but (3) is false. There are other members of which (3) and (2) are true; and (1) is not true *of us.*

We are in no doubt about the burdens we have taken on in staking our case against the Fallacy of Fallacies Schema. Certainly, there is no realistic prospect of doing so in the space of a single chapter. So we shall proceed as best we can, beginning with some issues we believe it necessary to explore in some detail before moving on to the negative and positive theses about *GOE.* This will leave us space enough to test these claims against only one class of fallacies, known collectively as "hasty generalization". The complete case against *GOE* is the business of *Seductions and Shortcuts.*

4.2 Logic's cognitive orientation

Since its inception 2,500 years go, logic has been thought of as a science of reasoning. Aristotle held that the logic of syllogisms is the theoretical core of the wholly general theory of argument called for in the *Topics.* Even centuries later, when logic took its momentous turn toward the mathematical, the idea persisted that the canons of logic regulate at least mathematical reasoning which, in some versions, is reasoning at its best. One of the striking features of mainstream mathematical logic is the distance at which it stands from the behaviour of real-world reasoning agents. In its anti-psychologicism, context-independency and agent-indifference, it is hardly surprising that mathematical logic endorses principles which real-life reasoners do not, and often cannot, conform to. Rather than taking this as outright condemnation of reasoning as it actually occurs, mathematical logicians have sought a degree of mitigation in the idea that real-life reasoning is correct to the degree to which it *approximates* to conformity to these ideal canons of strictness.[4] Although the

[4]Cf. Matthen (2002, 344), whose mention of it is disapproving: "Human reasoning *tries* to instantiate logic, but, because of the regrettable necessity of making do in the real world, it falls somewhat short. In this it is something like human virtue as Aristotle describes it—a second-best life imposed on us by the exigencies of the human condition." The more nearly correct view is that "[o]ur capacity for reason is dictated by symbolic complexity required for tasks other than truth maximization" (Matthen, 2002).

approximation-to-the-ideal view has had its critics (e.g., Gabbay and Woods, 2003a), other reactions have been more constructive and conciliatory. They are reactions linked together by the common purpose of extending and adapting mainstream logic itself, so as to produce systems capable of modeling aspects of actual reasoning which the standard systems of mathematical logic leave out of account. Within the logic community these extensions or adaptations include modal logics and their epistemic and deontic variations (von Wright, 1951; Hintikka, 1962; Kripke, 1963; Gabbay, 1976; Lenzen, 1978; Chellas, 1980; Hilpinen, 1981; Gochet and Gribomont, 2005), probabilistic and abductive logics (Magnani, 2001; Williamson, 2002; Gabbay and Woods, 2009), dynamic logics (Harel, 1979; van Benthem, 1996; Gochet, 2002), situation logics (Barwise and Perry, 1983), game-theoretic logics (Hintikka and Sandu, 1997), temporal and tense logics (Prior, 1967; van Benthem, 1983), time and action logics (Gabbay et al., 1994), systems of belief dynamics (Alchourron et al., 1985; Gabbay et al., 2002, 2004a, b) practical logics (Gabbay and Woods, 2003b, 2005), and various attempts to float the programme of informal logic.[5] Work of considerable interest has also arisen in the computer science, AI and cognitive psychology communities, with important developments in defeasible, non-monotonic and autoepistemic reasoning, and logic programming (Sandewall, 1972; Kowalski, 1979; McCarthy, 1980; Reiter, 1980; Moore, 1985; Pereira, 2002; Schlechta, 2004).

The net result of these considerable efforts is a marked reorientation of logic to the ins-and-outs of reasoning as it actually occurs. It may be said that, if in the aftermath of the mathematical turn it were ever in doubt, logic has now to some extent reclaimed its historical mission of probing how human reasoning does (and should) work.

This is a significant development. If logic is once more a science of reasoning, it is well to pause and take some note of what reasoning is *for.* It is clear upon inspection that, in a rough and ready way, reasoning serves as an aid to *belief-change* and *decision.* Certainly it seems true to say that is these aspects of reasoning in which the new logic (if we might appropriate that term) seems most to concentrate on (Gabbay and Woods, 2001b). This being so, an answer to our present question becomes apparent. Reasoning is an aid to cognition. Accordingly,

Proposition 2 (The new logic)**.** *Logic investigates reasoning in its role as an aid to cognition. Or, as we might now say the new logic is an investigation of* (requisite aspects of) *cognitive systems.*

[5]The informal logic movement comprises three over-lapping orientations. One is argumentation theory (Johnson, 1996, 2000; Freeman, 1991; Woods, 2003). Another is fallacy theory (Hamblin, 1970; Woods and Walton, 1989; Hansen and Pinto, 1995; Walton, 1995; Woods, 2004). Completing the trio is dialogue-logic (Hamblin, 1970; Barth and Krabbe, 1982; Hintikka, 1981; MacKenzie, 1990; Walton and Krabbe, 1995; Gabbay and Woods, 2001a, c).

4.3 Practical agency

A consideration of agency is central to our task. Our view of agency is set out in a PLCS—*a practical logic of cognitive systems,* which can be sketched as follows:

- A cognitively sensitive logic is a principled description of certain aspects of the behaviour of a cognitive system, chiefly of those aspects that figure centrally in *belief and decision dynamics*.
- A cognitive system is a triple of an *agent C*, cognitive *resources R* and *cognitive tasks J* performed in real *time t*.
- A cognitive agent is an information-processing device capable, among other things, of *belief, inference* and *decision*.
- A cognitive agent is always an agent of a certain type, depending on where he or it sits under a partial order that we will call "commanding greater (cognitive) resources than".
- Such resources include, but are not exhausted by, *information, time* and *computational capacity*.
- A cognitive agent is a *practical* agent to the extent that it ranks low in this ordering.
- Accordingly, *practical reasoning* is the reasoning of a practical agent.
- A cognitive agent is a *theoretical* agent to the extent that it sits high in this same ordering.
- Accordingly, *theoretical reasoning* is the reasoning done by theoretical agents.
- Practical agents include *individuals*.
- Theoretical agents include *institutions*.
- It cannot in general be supposed that practical and theoretical reasoning are geared to the same *goals* or *targets* and subject to the same *performance standards*.
- Compared with what theoretical agency can achieve, practical reasoner's operate with *fewer resources*.
- Compared with what theoretical agency can achieve, practical agents set *more modest cognitive targets*.

Accordingly,

Proposition 3 (Practical agency)**.** *Practical agency is triangulated by two main factors. One is the factor of comparative resource-scantness. The other is the factor of comparative target-modesty.*

We accept that ours is a somewhat unusual use of the word "theoretical". In the account given by *PLCS,* when an individual is, for example, trying to simplify a proof of the completeness of modal logic in time to meet an editor's deadline, he is engaged in practical reasoning, even though, in one standard sense of the word, the completeness problem is a theoretical problem. In putting the word to our uses here, we intend neither rivalry nor imperiousness. Ours is but another sense of the word, which we've introduced as a technical term. Even so, the gap between our use and other uses typified by the theoretical status of the completeness problem is not as large as one might think. There are legions of theoretical problems (in the completeness-problem sense) that demand the resources and epistemic standards that characterize theoretical agency in our sense. Most of NASA's scientific problems are theoretical in the completeness-problem sense, and NASA is an exemplar of theoretical agency in our sense. All the same, it is well to note that the word "practical" has no wholly natural (non-negative) antonym in English. So any candidate we might select is bound to strike the ear somewhat oddly.[6]

4.4 Cognitive economies

Seen in this way, practical agents operate in a *cognitive economy.* They seek to attain their targets with the resources at hand and with due regard for what they are naturally unfitted for. An individual agent's resources are for the most part available to him in low finite quantities. Given the multiplicity of his cognitive ambitions and the sundry demands of maintaining his balance in a world of constant change, there is an inevitable competition for the resources needed for the advancement of cognitive agendas. In much of what he does, an agent is a zero-sum consumer of his own resources. In lots of cases, he can also seek to draw down his competitors' resources as well. The zero-sum harshness of resource-draw demands that in most cases an agent pay attention to costs and benefits. This is not to say that his cognitive *targets* are economic (not usually anyhow) but rather that, whatever they chance to be, handling them rationally requires that these economic factors be taken into account. This is true of agents both practical and theoretical. Resources are finite for each and ambitions frequently outrun what resources are able to handle. The rationality of

[6]Various candidates have been proposed. We find that none generalizes in quite the desired way: *specialized, alethic* (or *doxastic*), *formal, precise, strict, context-free, abstract* and, of course, *theoretical* (in the completeness proof sense). For further discussion, see Gabbay and Woods (2003a, 13–14).

cognitive agency takes this factor of comparative resource-scantness into deep account. In virtually all that they do as cognitive beings, agents of both stripes must learn to economize.

Given these resource limitations, we may postulate for practical reasoners various scant-resource *compensation strategies.*[7] Leading the list, hardly surprisingly, is *the setting of targets of comparative modesty,* itself an instance of the adjustment of goals to the means available for their effective realization. Other strategies include:

- A propensity for *hasty generalization*
- A facility with *generic inference,* and other forms of non-universal generalization
- Ready discernment of *natural kinds*
- A propensity for *default reasoning*
- A capacity *to evade irrelevance*
- A disposition toward belief-update and discourse *economies,* such as reliance *ad verecundiam* upon the assurance of others
- A facility with *conjecture* (or, in plainer English, *guessing*)[8]
- A talent for *risk aversion*
- An architecture for *inconscious* or *implicit* cognition.[9]

We emphasize that scantness of resources is a comparative matter. By and large individual agents have fewer of them than institutional agents such as NASA or MI5. It is sometimes the case, though not uniquely or invariably, that resource-paucity makes for resource-*scarcity.* But it would be quite wrong to leave the suggestion that individual agents are resource-strapped by definition, as it were.

There are two quite general attributes that are unique to the practical agent, and which give him a clear advantage in the cognitive economy. One is the *emotional make-up* of (human) practical agents—in particular their capacity to feel fear, which plays a pivotal role in risk-averse inference. The other is that, to a degree far greater than applies to institutional agents, practical agents are capable of a timely response to *feedback mechanisms.* This is standing occasion for the practical agent to correct damaging or potentially damaging

[7] See Gigerenzer and Selten (2001).
[8] See Peirce (1992, 1931–1958, 7.220).
[9] This on the analogy of implicit perception, concerning which see Rensink (2000).

errors before the harm they portend is done. It conduces toward what we might call "an efficiently corrigible fallibility". Institutional agents, on the other hand, are notorious for their feedback-laggardness. It is a laggardness that routinely compromises efficiency and often compromises correction.

It would appear that, on the face of it, the list of scarce-resource compensation strategies is rife with fallacy, what with its endorsement of hasty generalizations and *ad verecundiam* and reasoning. Should we not conclude, therefore, that practical agency and practical reasoning are intrinsically defective? It is the business of Part II to deal with this question, at least in part.

4.5 Cognitive targets

We should now say a word about cognitive targets.

Proposition 4 (Cognitive targets)**.** *A target T for an agent X is a cognitive target for him (or it) if and only if T is attainable only by way of a* cognitive state *of X.*

For example, if X wants to know whether Y will accompany him to the movies, his target is met when he knows that Y will accompany him to the movies. The desire to know whether is X 's target. X's knowledge—that enables X to hit the target. T, then, is a cognitive target for X.

Not all cognitive targets expressly embed the desire to know; that is, they are not always overt calls for knowledge. X may desire to make a decision between options O_1 and O_2. Upon discovery of some new information, X may now be in a state of knowledge in virtue of which he decides for O_1 rather than O_2. X's state of knowledge closed his decision-agenda. So his decisional target was a cognitive target in our sense. Perhaps it might be said that in his desire to decide between O_1 and O_2, X was implicitly calling for the knowledge that would enable him to turn the trick. There is little point in semantic wrangles over the purported equivalence between "wants to decide" and "wants knowledge that will enable a decision". A target is hit when X no longer has the desire or disposition in terms of which it was constituted in the first place. This can happen in one or other of two ways that can be regarded as cognitive. In one, X is in a state of knowledge that causes X's desire to be *satisfied* or his disposition to be *actualized.* In the other, X is in a state of knowledge that *kills* X's desire or *cancels* his disposition. In the one case, X may desire to know whether his companion will accompany him, to the movies and it may happen that in coming to know that his companion will indeed accompany him that his desire is fulfilled. It may also happen that X desires to know who is using Department copier for personal purposes, and on coming to know that there is some indication that the culprit is his brother, his desire may lapse and his enquiry may cease.

In what follows, we focus on the first kind of case. Accordingly,

Proposition 5 (Attainment)**.** *If T is a cognitive target, then T's attainment requires the satisfaction of the desire embodied in T (or the actualization of its embedded cognitive disposition).*

4.6 The logic of down–below

It is well to emphasize that this talk of cognitive desire is largely an expository device, as indeed is the idea of an agent's cognitive targets. Targets can be likened to *agendas,* to whose examination our (earlier) companion work, *Agenda Relevance,* devotes a number of pages (Gabbay and Woods, 2003a, 37–40). This is not the place to repeat that discussion in detail, but there is some advantage in touching briefly on a few of its principal claims. One is that agendas (hence targets too) need not be consciously held or set, and need not be attended by express recognition of the means of their attainment. Cognitive targets are better understood as cognitive dispositions to be in certain kinds of mental states. But this is much too general a description to capture them adequately. Any cognitive agent, structured in approximately the way we ourselves are, is at virtually all times causally primed to be in the states to which he (or it) is, then and there, susceptible. Suffice it here to say that something counts as a cognitive target when it is of a type that could be consciously held, openly desired and deliberately advanced upon. That targets need not be thus held and advanced is further indication of how much of our cognitive careers are set out and dealt with subconsciously and (probably) sublinguistically. A short way of saying this is that a good deal of human cognition occurs "down below".[10] Reasoning, like cognition itself, also occurs automatically, unconsciously, sublinguistically, hence "down below". But logic investigates reasoning in its role as an aid to cognition. If logic is to honour its pledge to reasoning, it must be prepared in turn to probe the reasoning of down below. Given the constraints, both ethical and mechanical, that inhibit the exposure of human subjects to the vicissitudes of the experimental method, the logician is left with little choice but to abduce and to analogize. Whereupon is surrendered the ancient conceit that logic is the most certain and epistemically privileged of the sciences.[11]

[10] Other characterizations that have been used to capture the idea of reasoning down below are: *unconscious, automatic, inattentive, involuntary, non-semantic* and *deep.* We note in passing the general inequivalence of these descriptors (Gabbay and Woods, 2003a, 37–40).

[11] The logic of down-below is very much in its infancy. But already various ideas of how it might go have started to stir rather attractively. For a connectionist approach, see Churchland (1989, 1995); a *RWR* (representation without rules) orientation is discussed in (Horgan-Tienson, 1999a, b) and (Guarini, 2001); offline anti-representationalism is discussed in (Wheeler, 2001); a semantic space orientation is developed by Bruza et al. (2004, 2006) and connectionist neural net approaches are to be found in (d'Avila Garcez et al., 2002; d'Avila Garcez and Lamb, 2004) and (Gabbay and Woods, 2005, Section 6.8). For a criticism of the idea that logic imposes universal constraints on rationality, see (Matthen, 2002). (Bermúdez, 2004) explores the cognitive wherewithal of young infants and animals.

4.7 Generic reasoning

The identification of a practical agent as someone (or something) that performs his (or its) cognitive tasks under conditions of resource-paucity in pursuit of comparatively modest cognitive targets is one that states a generic fact about practical agents. What is claimed is that it is *characteristic* of the cognitive actions of practical agents that they are performed under such conditions in relation to such targets. It would be a mistake to ignore the plain fact that there are specific cases in which practical agents complete a task without at all depleting the resources required for its wholly satisfactory transaction. Neither is it the case that, in his generic thrall to comparative resource-paucity, the practical agent is invariably at a disadvantage. Whether he is disadvantaged in this way, or not, depends on the cognitive goals it would be appropriate for him to set for himself and on the cognitive wherewithal available for achieving them.

Unlike the universally quantified conditional sentences that inductive logicians recognize as full-bore (or Hempelian) generalizations, generic generalizations (if the pleonasm might be forgiven) are sub-universal in their reach. There is a considerable difference here. The generalization, "For all x, if x is a tiger, x is four-legged", is *brittle.* It is overturned by a single true negative instance. But the generic claim, "Tigers are four-legged", is *elastic.* It can be true even in the face of true counterinstances.[12] This provides the practical agent with further occasion to economize. If he ventures the generic claim rather than the strictly universal claim, he can be wrong in particular without being wrong in general—a nice advantage. Generic generalizations are less precise than Hempelian generalizations; but what is lost in precision is made up for in elasticity. Genericity, in turn, hooks up with the concept of default.

Proposition 6 (Genericity and default inference)**.** *Given the generic claim that tigers are four-legged, together with the fact that Pussy is a tiger, the inference to "Pussy is four-legged" is an inference to a default. What makes it a default is precisely that "Pussy is a four-legged tiger" could be false without making it false that tigers are four-legged.*

Hasty generalization is intimately linked to genericity, which in turn is intimately linked to natural kinds. To see a tiger as the kind of thing it is involves having some grasp of properties it possesses as a thing of that kind. But this is knowing something about what is characteristic of tigers, hence true of them by and large. Seeing that Pussy is a tiger—that Pussy is of the tiger kind, rather than, say, of the James Bond villainess kind—involves an appreciation of what things of that kind are *like;* that things of that kind are by and large four-legged,

[12] For genericity Carlson and Pelletier (1995) is essential reading.

for example. It is doubtless an over-simplification, but something like the following holds true: that appreciating that this thing is of the tiger kind involves appreciating that various of this thing's properties are by and large properties of all things of that kind. So natural kind recognition involves hasty (generic) generalization of kind-properties.

The distinctive advantage of generic generalizations is that they can be retained without qualification even in the face of known counterinstances.

Hempelian generalizations are disabled by true counterinstances, and require, if not outright abandonment, nothing less than reformulation. There are four basic ways of achieving such reformulations, each problematic. One is to hit upon a principled means of exclusion, that is, a means that serves to exclude the requisite class of the true counterinstances that is stateable without making specific mention of them. Another is to restate the original generalization and append to it, one by one, classes of known exceptions. A virtue of the first approach is that it avoids the ad hocness of the second. A drawback is that it is often unknown as to what constitutes, with appropriate generality, the qualification that transforms a defeated generalization into a live one. Attesting to this difficulty is the liberal invocation of *ceteris paribus* considerations. A dubious evasion if ever there were one, retention of the original generalization is made possible only by the expedient of "paying in advance" for unspecified counterexamples. A fourth remedy is the hoary old device of approximation, in which a generalization, though defeated by counterinstance, is retained as approximately true.

Let us consider these in order, beginning with the base case.

- All tigers are four-legged.

Option one provides for something like

- All *properly made* tigers are four-legged.

This is troublesome. If "properly made" here entails "four-legged", the revision is vacuous. If it doesn't entail "four-legged", it is simply useless as things stand how "properly made" achieves the desired exclusions. Of course, various unpackings are possible. We might be invited to consider that properly made tigers are those with the wherewithal to preserve four-leggedness in the descendent class of tigers; but unfortunately this presupposes that all tigers (now) are four-legged, or that one or other of the very reformulations currently under review holds true of them, taking us again too close to circularity for comfort. But circularity aside, the present means of saving this low-order generalization also involves a considerable, and unwelcome, complexity.

A further option gives us

- All tigers, *except those with certain kinds of congenital effects or those injured in certain ways,* are four-legged.

This is also problematic. The trouble is the unspecificity of "certain kinds of" and "certain ways". Left unspecified, there is reason to doubt the generalization's truth. But if the intended specificity is presumed, the generalization is vacuously true. One way of achieving the exceptions without running foul of these difficulties is to list the exceptions, one by one, as in

- All tigers are four-legged *except Pussy, Fred, Baby and Monster.*

But this is hopeless. No one wanting to assert the generalization safely has the foggiest idea as to how the completed list goes.

The *ceteribus paribus* option gives us

- *Other things being equal,* all tigers are four-legged.

Here, too, the unspecificity of "other things being equal" threatens to falsify the generalization, and its specificity threatens to trivialize it. The same is true of

- It is approximately the case that all tigers are four-legged.

If "approximately" means "except those that aren't", we have triviality. If it means something less specific, it cannot be ruled out that it imposes the wrong qualification. It would be a mistake to leave the impression that this brief review of the options is decisive against the reformulation view of defeated Hempelian generalizations. But enough has been said to indicate how difficult and complex such repairs must prove to be. In plain English,

Proposition 7 (The economic advantage of genericity)**.** *Defeated Hempelian generalizations are* hard to fix. *Generic claims with true negative instances* don't have to be fixed.

4.8 Epistemology

Apart from its role in investigating reasoning in its role as an aid to cognition, logic has always carried epistemological presuppositions. Even in the comparatively small historical space of the century just past, one sees the passage from the apriorist, foundational, Platonized realism of Frege and Russell to the pragmatism of Quine, with a concomitant explosion of logical pluralism.[13] But once logic re-adopted *agents* as a central theoretical parameter, it became necessary to pay some degree of attention to what agents are like, to

[13]It may be more accurate to characterize Frege's realism as more Kantian than Platonic. Certainly Frege is not a realist about sets ("courses of values") in the way that Gödel is. Also, it must be acknowledged that as early as 1907 Russell on occasion was quite openly a pragmatist about the justification of "recondite" principles of logic. Strangely, this would later be a position taken up by Gödel. Concerning the first point we are indebted to Ori Simchen for helpful suggestions. Concerning the second, see Irvine (1989). Rodych examines whether Gödel's Platonic ontology is reconcilable with his pragmatic epistemology.

what their interests are and what they are capable of. If logic is to deal with reasoning that advances (or retards) an agent's cognitive agenda, it is necessary that it take note of what the agenda is and how it relates to the agent's wherewithal for advancing it. Any such observation will be incomplete until it is buttressed by an appreciation of the general conditions under which an agent achieves epistemic fulfillment.

If we re-examine various of the conceptual skeins of the new logic, especially in its emphases on defeasibility, non-monotonicity and defaultedness, it can be seen that at present the dominant epistemological presumption is *fallibilism.* Fallibilism is expressly endorsed in the present authors' multi-volume work, *A Practical Logic of Cognitive Systems.*[14] In the present chapter we re-establish that commitment. The idea that real-life cognizers are fallible agents has a certain clear attraction. It expressly embeds the idea of *error* or *mistake,* surely not an irrelevant circumstance for anyone writing about *fallacies.*

4.9 Fallibilism

Fallibilism is a philosophical thesis about error. Since fallacies are errors, it might well be expected that the philosophical thesis that fallibilism is would afford us some insight into the kind of error that fallacy is. Needless to say, the fruitfulness of the connection cannot be guaranteed in advance. It may turn out that there is less to it than we might have supposed. It cannot even be ruled out that there is nothing to it. But if that were so, it would be very odd; it would call out for an explanation.

In its most interesting form, fallibilism is a normative claim. It holds that

Proposition 8 (Fallibilism)**.**

(i) Not only do actual agents sometimes make errors; but

(ii) even when operating at optimal levels occasional error is unavoidable; and yet

(iii) it is wholly rational for a real-world cognitive agent to deploy cognitive strategies (including the adoption of rules of inference) that he (or it) knows in advance will on occasion lead him (or it) into error.

Examples abound. Deductive rules can lead us to false conclusions; inductive strategies can induce the acceptance of defective generalizations; abductive reasoning embodies the risk that attends conjecture; and on and on.

Clause (iii) encompasses two quite distinct notions of error; it is important to give each its due. To mark this difference it helps to take note of another one. It is the contrast between

[14] Of which volume 1 is Gabbay and Woods (2003a) and volume 2 is Gabbay and Woods (2005). Additional volumes will appear in due course.

(a) *Error-elimination* strategies

and

(b) *Error-susceptible* strategies

A good example of an error-susceptible strategy is a default inference from generic premisses. As we have said, a generic claim is a form of general proposition that remains true in the face of (certain classes) of true negative instances. Since a default is a conclusion of an inference in which the "major" premiss is generic, it imbibes this same feature, but in a particular way. Though some classes of negative instances of a true generic claim, *F*s *G*, are also true, it is *not an error* to claim that *F*s *G*, and it is not inconsistent to say that although some *F*s don't *G*, *F*s nevertheless *G*. But given that *F*s do indeed *G* and that *this* is an *F*, we have it as a default that this *G*s. The genericity of "*F*s *G*" allows that "This *F G*s" is false. If so, then the default that is our conclusion in this case *is* an error. This is important. Although, as we have it here, the premisses of the default inference are error-free, and the inference in question is correct, the inference is not of a kind as to preserve freedom from error. So in the absence of information to the contrary,

Proposition 9 (Default inference)**.** *It is reasonable to infer a default from a set of premisses, of which the major is a generic claim and the default an instance of it, notwithstanding that such inferences are not error-avoidance preserving, and that the reasoner is aware of this.*

4.10 Errors of logic

Standard approaches to deductive and inductive logic are wholly concerned with error-elimination strategies. If, as in the case of deductive logic, the error to avoid is invalidity,[15] that error *is* voided whenever the deductive protocols are applied properly. If, as in the case of inductive logic, the error to avoid is inductive weakness, that error *is* avoided whenever the probability rules are applied properly. This carries the suggestion that no such error is possible for any agent who deploys the requisite protocols correctly. Category (b) is different. Its protocols include those for generic inference, as well as various procedures for presumptive and default reasoning. Even if perfectly applied it cannot be guaranteed that they will hit their respective targets. They are, therefore, error-susceptible protocols. This bears on fallibilism in a twofold way. It provides that

(iv) Actual agents are prone (and know it) to applying both error-elimination and error-susceptible strategies incorrectly.

[15] For ease of exposition, we allow invalidity to stand in for the others: inconsistency and logical falsehood.

and it reminds us that

(v) It is insufficient for the cognitive agendas that agents actually have to deploy only strategies of type (a).

Accordingly, not only are actual agents destined to make *application errors,* they are also drawn to the use of strategies whose entirely correct application embodies the occasion of error; in other words, they are also prone to *susceptibility errors.*

It lies at the heart of the present conception of fallibilism that errors cannot be simply "wrong answers". In an extended sense, this is precisely the view that prevails in the error-avoidance precincts of standard logic. It allows us to characterize an argument (i.e., a sequence of propositions) as erroneous simply when it fails to be valid. It allows us to characterize an argument as erroneous simply when it fails to achieve a certain degree of inductive strength. This is plainly not the sense of error that fallibilism seeks to make something of, for then a considerable abundance of perfectly reasonable inferences would have to be classified as errors. What makes this so is that the great percentage of reasonable inferences actually drawn by real-life agents are neither valid (in the sense of deductive logic) nor inductively strong (in the sense of the calculus of probability).

What these conceptions of error lack is an aspect central to the fallibilist approach to the matter. It is the factor of *illusion, inapparency* or *agent-unawareness.* Accordingly,

Proposition 10 (Inapparency)**.** *It is fundamental to the conception of error that an error is a failure or a defect of which its committor is unaware.*

This, to be sure, is the common meaning of the term, as with its near-synonym "mistake". It is a conception that might well irritate those who believe that logic has no business investigating states of mind, but it can hardly be refused by a logic in which a central parameter is the real-life agent. Real-life agents come equipped with states of mind, like it or not. The idea of error as inapparent defectiveness is as old as logic itself. Aristotle expressly advances the notion in *On Sophistical Refutations.* He called them *fallacies.*

Aristotle held that the most general thing to be said about a fallacy is that it is an argument that appears to have a certain property which in fact does not have it. In *On Sophistical Refutations,* Aristotle was more narrowly focused. He wanted to characterize a certain kind of argument in which the notion of syllogism plays an integral role. Aristotle defined a *refutation* as a syllogism whose conclusion contradicts an opponent's thesis and whose premisses are drawn exclusively from the opponent's own concessions. Accordingly, a *sophistical* refutation is an argument thus conceived that seems to be a syllogism but it isn't.

In one of his first tasks as a logician, the founder of logic draws our attention to this phenomenon of false inapparency. In one place, he tells us that it is "the death of argument" (Woods, 2004, Prologue). *On Sophistical Refutations* takes up the task of classifying these bad arguments. Aristotle's list runs to thirteen, though there is reason to believe that he didn't think this an exhaustive inventory. Many pages of this little treatise are given over to brief examinations of where the fault of these bad arguments precisely lies. But no one, least of all Aristotle, thinks that these diagnoses are complete.

It is well to note that in *On Sophistical Refutations* comes close to sharing an assumption with modern formal logic. This is the assumption that the notion of error that these logics adumbrate is one of *deductive insufficiency.* In the case of modern logic, it is invalidity pure and simple. In the case of Aristotle, it is *either* invalidity pure and simple *or* the failure of one or other of the further conditions that Aristotle places on syllogisms. In other words, it is the error of syllogistic invalidity. When one tests this nearly-enough common assumption against actual argumentative practice, it is easy to see that there is something wrong with it. Taking modern logic as our example (it easily extends to fit the syllogistic case), it is no secret that validity is hardly ever an agent's cognitive target. Even in those relatively isolated instances in which a logician wants to know *whether* an argument is valid, producing an argument that *is* valid is neither necessary nor sufficient for the attainment of that target. To illustrate:

1. If an agent X wants to know whether $\langle\{P_1, \ldots, P_n\}, Q\rangle$ is valid, then producing the valid argument $\langle\{P\}, P\rangle$ doesn't hit that target.

2. Neither is it hit just by producing the *very argument* $\langle\{P_1, \ldots, P_n\}, Q\rangle$ (assuming it to be valid); for X may not know that it is valid.

3. X might hit the target by checking the Answers section in a logic textbook. But then he hasn't himself produced anything that is valid, and the answer itself might well consist of the single word "Valid".

Beyond these comparatively rare cases, an agent's cognitive target is not aimed at validity, even though validity may be the requisite *standard* that the attainment of that target may require. If an agent desires a proof of a proposition he will fail unless his reasoning meets the requisite standards, of which validity is one. Clearly, then, one's cognitive target might well be such that it will not be attained unless the validity standard is met. But it is misleading to say that validity is itself the agent's target.

Although the validity standard is sometimes necessary for target attainment, most cognitive targets neither require nor are advanced by fulfillment of the validity standard. We have it then, that invalidity is not, just so, an error, notwithstanding our assumption paragraphs ago that if modern logic had a concept of

error, it could only be invalidity. Invalidity is an error only in relation to cognitive targets for whose attainment the validity standard applies. In so saying, it may occur to us that this is not, in fact, contradicted by the presumptions of modern logic. Whatever its targets, mainstream deductive logic makes it a condition of attainment that the validity standard be met. If this is so, it is largely implicit. It is not much talked about by logicians.

Suppose that we were satisfied with the suggestion that the targets that (however tacitly) call for deductive reasoning require that the validity standard be met. This would be a good place to call attention to an impressive omission.

Proposition 11 (Accounting for error)**.** *Standard deductive logics embed a notion of error, but no such system gives an account of it.*

Why would this be so? Two reasons stand out. One is that, in its subscription to formal languages, standard systems of deductive logic seek to eliminate the linguistic confusions that give rise to fallacies (Frege, 1879; Peirce, 1992; Tarski, 1956; Quine, 1970). The other is that, in as much as deductive logic lacks the capacity to produce a formal theory of invalidity for natural languages, it may be thought that the concept of error lies beyond logic's theoretical embrace (Johnson, 1967; Massey, 1981). We take up these issues in (Gabbay and Woods, 2009).

Targets carry standards for their attainment. Something is an error if it fails to meet the required standard. Again, not just any valid argument will meet the validity standard of every cognitive target that embeds a validity standard. Speaking this way relativizes standards to targets and imposes the same relativity on the concept of error. One can only wonder whether these things might not be subject to further relativities. The answer is that they are.

An agent might wish to know the proof of the completeness of formal arithmetic. If so, he would have made an error. His target is defective in a quite particular way; it embodies a false presupposition. An agent might set himself the target of acquiring a Ph.D. in quantum computation. But if he is 92 years of age, a high school drop-out, and possessed of a modest I.Q., he too has made a mistake. It is not that the target of getting a Ph.D. in quantum computation is impossible to attain, but rather that it is impossible *for him* to attain. It was the wrong thing to aim for, given this agent's cognitive resources. Here, then, is another pair of factors that bear on the issue of error.

Proposition 12 (Error relativity)**.** *Something may be an error in relation to the standards required for target attainment, in relation to the legitimacy of the target itself, or in relation to the agent's cognitive wherewithal for attaining it.*

4.11 Parameters of the subpar

Let us tarry awhile with this idea of subpar cognitive performance. So again we ask: What is it to judge that someone's cognitive conduct is not up to snuff? It is to find fault with the action in the light of various criterial considerations. As we saw, one is what the agent's target is. Another is the standard that he needs to hit for that goal to be attained; in other words, the agent's means to that end. A third factor in judging an agent's cognitive performance is his *general competence.* In mentioning it, we reveal an interest in determining whether this is a goal whose satisfaction by hitting that standard is something that he is able to do. A fourth consideration has to do with *collateral considerations.* An agent may have the general capacity to achieve a certain goal in a certain way, but, owing to present particularities, not be able to achieve it or to achieve it in that way. In citing this factor, we are recognizing the importance, beyond general competence and means-end adroitness, of cognitive resource-contingencies such as (again) *information, time* and *computational capacity.*

Jointly, these factors give a blueprint of an agent's performance of a cognitive task. A cognitive target T is either attainable or not. (A proof of Fermat's Last Theorem is attainable; a proof of the joint consistency and completeness of Peano-arithmetic is not.) If a goal is attainable, then for any agent X, it falls within X's general competence or not. (A proof of the completeness of modal logic was within Ruth Barcan's reach but not, we may suppose, Hannah Arendt's.) If X has an attainable goal that lies within her general competence, the means she selects (or the standard she sets) may be appropriate for that goal or not. If, for example, X undertakes to show for some proposition P that P is something that might reasonably be believed, her standard may include an argument for P that meets the standard of validity. In her quest to justify a belief in P in this way, X would be at risk for two performance errors. Either validity may be an inappropriate way of achieving this goal, or it may be appropriate but beyond X's reach. X might not know how to construct valid arguments (perhaps she is a struggling student of First Year logic). If X has an attainable goal that is within her general competence, for which an appropriate means S is also within her grasp, X may lack additional resources R necessary for the completion of her task. She might not have information enough to command the desired means; or she may lack the time to achieve her objective in this way; or she may lack the computational power to do the calculations that her task requires of her. Alternatively, given the comparative scantness of such resources for real-life individuals in actual situations of cognitive effort, an agent may simply lack the means of achieving the goal. If, again, an agent's goal is to show that it is reasonable to believe that P, she may decide that an axiomatic proof of P is not a means for which she is adequately resourced at present; and she might try instead for a conditional proof relative to what is

widely held by experts (wherewith the potential for *ad verecundiam* error may present itself).

We may say, then, that

Proposition 13 (Further relativities)**.** *There are several basic ways in which an agent X's cognitive performance can go wrong:*

1. *X might set himself a simply unattainable target T.*
2. *X might set himself an attainable T that is not within his general competence.*
3. *X may set himself an attainable T for which he is generally competent, but his selected means (or goal-realization standard) S is either beyond his reach or inappropriate to the task at hand.*
4. *X may set himself an attainable T for which he is generally competent and set himself an appropriate S that lies within his reach, and yet he might lack necessary collateral resources R.*

When this last condition is met, we shall say that T is an attainable goal for which X is generally competent, that S is a realizable and appropriate means for X to set in relation to T, but that for lack of such things as information, time and fire-power, T sets a task that is *too big* for X.

Ed Koch, on his walking tours of New York when he was mayor, famously would ask, "How am I doing?" We daresay in inviting this assessment of his performance as chief magistrate, he was unaware of all the details of the template that structures a fair response. It is a template that calls for the assessment in terms of T, S, R. These are the structural elements necessary for a finding of "subpar" with respect to the ranges of cognitive performance that draw the attention of fallacy theorists. They apply to Ed Koch. And they apply to the rest of us as well.

4.12 Ought and can

No practical agent can be faulted for mismanaging a cognitive task that is too big for him, although he might well bear some responsibility for having acquiesced to such a task. Whatever we say about such (mis)performances, they are not fallacious. In *some* sense, a principle of "ought"-implies-"can" is at work here. There is, however, a certain confusion that we should try to avoid. In saying that a better performance is not possible for agents of type so-and-so, it is not always required that we deny its betterness. It is required only that we resist the inference that a possible performance that is less than better for agents of this type is *subpar for them.*

There is in these reflections occasion to consider a sister principle to "ought"-implies-"can". We could call it "can"-doesn't-imply-"ought"; it has

the virtue of being in general even more obviously true than its kin. It is not, however, trivially or vacuously true; for especially in enquiries into human cognitive performance, exceptions to it are expressly countenanced, some having the status of scientific postulates. In any account of human practice in which optimization is held to trump satisficization, and it is also assumed that it is always better to do one's best, that "can"-doesn't-imply-"ought" is conspicuously disregarded. Variations of its opposite, "can"-implies-"ought" flourish in standard accounts of belief dynamics and rational decision-making (Alchourron et al., 1985; Raiffa, 1968). However, it is well-attested in actual practice that practical reasoners often saticifice rather than optimize, even when optimization is available to them as an achievable goal. In such practice there is an important reciprocity between targets and standards. What a cognizer needs to know and how he sets about to know it is a matter of what the knowledge is wanted for. Peirce once quipped that we know who our parents are by hearsay. Given the documentary thoroughness of modern life, to say nothing of the identificatory capacities of DNA technology, one could know more of one's parentage—and know it more strictly—than the run-of the-mill offspring has (as the saying goes) "time for". It is not that this larger and more strictly realized knowledge exceeds his reach. In the general case it exceeds his cognitive goal (to know whom to call "Mum" and "Dad") and imposes a cognitive standard that he has no need of. For ranges of cases, "can" clearly does *not* imply "ought". When an agent pursues a target or a standard, or both, that is bigger than it need be, we shall say that their pursuit by that agent is a case of *overkill.*

Before leaving the suggestion that a version of "ought"-implies-"can" holds for the assessment of cognitive performance, care needs to be taken not to trample on the latitude underwritten by fallibilism. If fallibilism provides that there are cognitive procedures that it is rational to execute even in the knowledge that they are virtually certain to lead one to occasional error, and if it also holds that there is a sense in which such procedures can't be abandoned, then fallibilism allows for a conception of error that a reasoner can't help committing or can't help committing without cost to his procedural rationality. So we must not allow the sense in which "ought" implies "can" to trespass on this provision.

Consider now a real-life individual who has set himself the task of advancing his cognitive agendas—of living his cognitive life—on the model of NASA. Given his resources and the loftiness of his cognitive ambitions, his cognitive life is a guaranteed disaster. Cognizing on the model of NASA is too big a task for any individual. In one sense, it is quite right to "forgive" *X* his cognitive failures. One can't be expected to achieve what one hasn't the means of achieving. Even so, *X* didn't have to set his targets so high. It was well within his power to select his targets with a view to his ability to meet them. If this is

so, his massive failures are subject to disapproval of higher order. They were the inevitable outcome of unrealistic targets that he needn't (and shouldn't) have pledged to.

Be that as it may be, there still remains the utterly central question of whether, and to what extent, an agent—any agent of whatever type—can be held responsible for an *error,* given that an error is something that he cannot, then and there, see as such. Take a case. Let *X* now set his targets more realistically. Let us say that they are of a type for which he has the requisite competence and the necessary resources. They are not too big for him. Even so, we have it by the very idea of error that if *X* errs in his quest to attain *T*, his error is something inapparent to him. And we have it by the meaning of fallibilism that the best that is in *X* rationally to do involves him in cognitive procedures that will on occasion expose him to error, that *X* knows this; and that knowing it is no affront to his reasonableness in retaining those very procedures. Against this, there is a strong disposition to find fault with at least those errors that have acquired membership in *GOE.* As Douglas Walton has it, attributing such an error to *X* is one of the harshest criticisms that can be leveled at *X*'s performance (Walton, 1995). The literature also embeds the widely-held view that fallacies are errors of a kind made avoidable by due care. But, as we see, neither of these views rests well with any view on which errors are undetectable, especially when such a view is embedded in a fallibilist epistemology.

4.13 Inapparency

On the face of it, a theory of fallacy has a twofold task. Since a fallacy is an error, a theory of fallaciousness should embed an account of error. Since a fallacy is an inapparent error, a theory of fallaciousness should contain an account of the factor of inapparency. There is, to be sure, an element of redundancy in putting it this way, since inapparency is intrinsic to error. Accordingly, a theory of error would also have the task of dealing with inapparency. But there is no harm in listing the inapparency requirement as a separate theoretical responsibility, if only to lend it an emphasis to which the literature is largely inattentive.

Inapparency, then, is intrinsic to error. In committing an error, there is something its committor has *over-looked,* something that he has failed to *see.* It bears on this that in its most common meaning a fallacy is a "common misconception", a belief which, although false, is widely and confidently held. It is an attractive belief whose falsity has escaped the committor's attention. The psychological literature draws a useful distinction between *performance* and *competence* errors. A performance error arises from contingent factors such as fatigue, intoxication or intention. Competence errors spring from more structurally embedded kinds of inability. If a good night's sleep might arouse a

reasoner from yesterday's performance errors, it will do him no good on the score of incompetence. These are transgressions whose avoidance exceeds the very design of the committor's cognitive wherewithal. A particularly good example of a competence error is one that arises in the treatment of a problem whose solution requires an effort that exceeds the computational capacity of the type of agent in question. Competence errors are not, however, a particularly good example of the sort of inapparent misstep we are currently discussing. The reason for this is that

Proposition 14 (Abiding competence)**.** *It is a compensation strategy among beings like us to tend to avoid the employment of cognitive protocols that exceed their competence.*

A case in point: An exhaustive check of our present web of belief for truth functional consistency would involve us in a computational explosion vastly beyond the reach of what we are built for. But there isn't the slightest empirical evidence that, when beings like us *do* attempt to reconcile their beliefs to some standard of consistency, this involves anything like even an exhaustive search.

A further locus of inapparency has been held to be the argument (or piece of reasoning) itself. So seen, an argument (or inference) that we erroneously pledge to (or erroneously draw) is one whose defectiveness is inapparent even to a well-rested and competent cognitive agent, arising from a kind of camouflage or disguise. Needless to say, these are rather anthropomorphic metaphors, having a more literal application in cases of an interlocutor's intention to deceive his opponent. But the factor of disguise is, on this view, lodged not in the committor's malign intention but rather in his warp and woof of argument or inference. Of the many theorists who subscribe to such a view, perhaps it is Lawrence Powers who puts the point most clearly:

Proposition 15 (Powers' inapparency principle)**.** *The false inapparency of an erroneous argument or inference is an* objective feature *of the argument or inference, rather than an* interactive feature *of them with a cognitive agent (Powers, 1995).*

We leave it to Powers to identify those objective features. We ourselves are minded to look elsewhere—to the very structure of cognition itself—for an especially important and, in its way, objective, locus of false inapparency.

Let us observe that in one of its most common meanings the word "believe" (and its cognates) admits of a striking first-third person asymmetry. On this usage, when *Y* says of *X* that *X believes* that *p*, *X* would say of *himself* that *p*. Legions of philosophers have been right to observe that *self*-ascription of belief constitute a kind of attenuated or qualified subscription to the proposition at hand. But in the present meaning of the term, the *other*-ascriptions of a belief that *p* leave it entirely open that the person to whom the belief is attributed

holds to p (and is right to) assertively and without qualification. Accordingly, for the sense of "believes" in question,

Proposition 16 (Belief as knowledge-claim)**.** *Whenever it is true for Y to say of X that X believes that p, it is true that X takes himself as* knowing *p* to be true.

Proposition 16 is a blindspot context (Sorensen, 1988). Whenever it is true for Y to say of X that X believes that p, then for X to say of himself

- p, and I believe p

would constitute a *blind-spot.* That is to say, in the absence of further information, any person to whom the bulleted admission were directed would lack the means to ascertain just what the utterer's epistemic state toward p has been claimed to be. Is the utterer saying that he knows that p? Or is he saying that he (merely) thinks that p?

Consider now an agent X's cognitive target K. Suppose that K is such as to be attainable only when X is in an epistemic state k. Let k be the state in which it is true to say that X knows that p. X's target T is occasion of a kind of *cognitive irritation.*[16] X is so constituted and so related to T that he aspires to be in a state in which the irritation is relieved.[17] We have known at least since the presocratics that although being in k is the state that X is required to be in for T to be attained, it is *not* required for X's cognitive irritation to be relieved. Irritation-relief is one thing. Cognitive attainment is another. From the third-person perspective, this is not a difficult contrast to command. But from the first-person perspective, it is a contrast that collapses, and is recoverable if at all only in the person's own reflective aftermath. When that reflective aftermath is at hand, the first-person can now say what the third-person could have said all along: X only believed that p, rather than knowing it. We have it, then, that when X is in a state of belief that relieves the cognitive irritation occasioned by T, he is in a state which he takes to constitute attainment of T. Not only is that state, b, *not* the same as k, but X's being in k, carries no phenomenological markers over and above those carried by b. Accordingly,

Proposition 17 (Phenomenologically structured inapparency)**.** *By the phenomenological structure of individual cognitive agency, the difference between being in b and being in k is phenomenologically inapparent. So where one indeed is not in k, being in b disguises that fact.*

[16] We must take care with the metaphor of irritation. Not every irritation of the human system that is put right by the requisite causal adjustments is something the human agent is either conscious of or openly desirous of remedying. Given that cognition can be so deeply implicit, we require the same latitude be extended to the idea of cognitive irritants.

[17] Such aspirations flow from what St. Augustine calls "the *eros* of the mind". In Gabbay and Woods (2005) it is called "cognitive yearning".

If this is right, then the capacity for, indeed the likelihood of, false apparency is structured by the phenomenology of cognitive states. For one thing, it seems not so much to be a property of a given argument or a given piece of reasoning, but rather a factor intrinsic to the possession of *b*-states in relation to *T*'s that call for attainment by way of *k*-states. It bears repeating that cognitive *relief* is not, just so, cognitive *attainment;* it is rather the appearance of it. Certainly in our disposition to confuse relief with attainment, there need be not the slightest hint of fatigue or intoxication. In other words, our present confusion seems not to be, or to arise from, performance errors. Given that such confusions appear to be intrinsic to the phenomenological structure of cognitive states, it lies more in the ambit of the competence error, hence reflective of an objective fact about how individually cognitive agents are constructed.

4.14 Valuing validity and inductive strength

Let there be no doubt, when truth-preservation is indeed an agent's cognitive target, validity is a necessary part of the standard for its attainment. However since truth-preservation does not, just so, guarantee the proof of anything,[18] truth-preservation rarely achieves the status of cognitive target, and rightly. In realistic settings, truth-preservation is itself valued not as a target but as a standard. In other words, in realistic settings, truth-preservation and validity are the *same* standard.

Valuable though it is in some settings, it is easy to think too much of validity; at least this is so when validity is monotonic. Let T be a target that calls for a valid argument. Let V be such an argument. Let K be a proposition that contradicts V's conclusion and is not in V's premiss-set. Let us also put it that the discovery of K is a huge surprise for X. Let V^* arise from V by addition of K as premiss. Since V is valid, so is V^*. But it is clear that although V^* is valid, it is not of the slightest use to X. It is not of the slightest use notwithstanding that it is a valid argument retaining all the premisses of V, which, until the discovery of K, we may suppose to have been of considerable use to X. For it was a valid argument none of whose premisses is a proposition that X then had any reason to doubt. What we see, then, is that *validity-preservation* is not a realistic standard even for targets for which validity is a necessary standard.

Validity is unresponsive to new information. In this respect, it is natural to suppose that *inductive strength* is the more useful standard. Its usefulness is a matter of its *non*-monotonicity. Its non-monotonicity makes it responsive to new information. This is true but not especially availing. Let I be an argument whose conclusion C has a requisite degree of conditional probability given its

[18] Save for the corresponding conditional of the argument that the target's attainment standard requires to be valid.

premisses $P_1, \ldots, P_n$. I is an inductively strong argument. Suppose now that K is new information that falsifies C. Since K is new, it is not in I's premiss-set. Let I^* arise from I by addition of K as premiss. Notwithstanding that I is inductively strong, I^* is inductively impotent. It is clear that, even though new information can collapse inductive strength, there is an inductively strong argument available to X that is wholly untouched by the new information. This is argument I, and the reason that it is wholly untouched by K is that K is not in its premiss-set. What this tells us is that, even where inductive strength is part of a target's attainment standard, it is a smaller part than might have been supposed. As we now see, validity-*preservation* is not part of the standard of any target whose attainment calls for validity. The reason for this is that validity provides it automatically. Validity-preservation is a free-rider. But with induction we may say that the reverse is true. That is to say, given any target for which inductive strength is part of the attainment standard, preservation of inductive strength in the face of new information is also a requirement. It is easy to see that this latter imposes on an agent's inductive targets the weightier requirement that the inductions be made from up-to-date information, i.e., that they not admit any information that collapses inductive strength. In the inductive cases, falsifying new information matters *inductively.* In the deductive cases, falsifying new information does not matter *deductively.* In both cases, however, what matters more is the *state of the information* from which conclusions are drawn.

PART II: FALLACIOUS COGNITIVE VIRTUES

This would be a good place to restate our principal theses about the fallacies.

Proposition 18 (The no-fallacy thesis)**.** *Not all of the Gang of Eighteen are fallacies. Those that are are not characteristically committed by beings like us.*

Proposition 19 (The cognitive virtue thesis)**.** *Several of the Gang of Eighteen are cognitively virtuous scant-resource compensation strategies.*

In what remains of this essay, we shall attempt to vindicate these claims as they apply to hasty generalization.

Limiting the defence to just one might well strike the reader as favouring our cause with an artificially small sample. But the reason is the want of space.

4.15 Hasty generalization

Hasty generalization, also known as *thin-slicing*,[19] is an error when committed in response to a cognitive target T, whose attainment embeds the standard S of inductive strength. For example, T might be the goal of reaching a

[19] See Ambady and Rosenthal (1993) and Carrere and Gottman (1999).

generalization about some subject with scientific accuracy. In that case, it is reasonable to require that his (or its) reasoning rise to the standard of inductive strength. It is easy to see that it is comparatively rare for individual agents to set targets of such loftiness. If an agent is part of a drug assessment team for Health Canada, we would certainly expect him and his colleagues to set themselves such a T and bind themselves to such an S. But an indication of how comparatively rare this, even for this individual, is the comparatively generous command he enjoys of Health Canada's resources for T—time, information, computational power, money, infrastructural and cultural encouragement, and so on. To the extent that this is so, this person and his mates are not acting as practical agents. They have teamed together and they have attracted levels of support in ways that give their efforts the kick of theoretical agency. Most practical agents lack the rudiments of scientific method, whether knowledge of how to compose a stratified random sample or of how to calculate even low-level conditional probabilities. What is more, if they did know, it would in very large ranges of cases be beyond what they had either time or computational capacity for (Harman, 1986). There is a widely received view that all of this is true but beside the point. For even practical agents (it is said), limitations and all, are performing at their ampliative best when they strain against these limits and approximate to the behaviour called for by the methods of science.

This, of course, is *scientism.* Saying so doesn't take us much beyond name-calling. So something further must be said against the view that in matters ampliative it is best to conform one's reasoning to the requirements of induction. In preceding sections, we have given out part of what we take to be the correct treatment of hasty generalization. We have seen that when one generalizes hastily, one often generalizes to a generic proposition rather than to a universally quantified conditional proposition (full-bore Hempelian generalizations, as we called them). One of the chief virtues of proceeding in this way is that even when as instantiated default is false, it is necessary to forgo the instantiation but not to repair the generic generalization whence it sprang. There is a considerable economy in this, needless to say; and that alone vests it with an attractive advantage. A further point of importance—perhaps the fact of dominating significance here—is that even when we seek the lofty goals of scientifically pure induction, we tend to generalize hastily. In beings like us, hasty generalization is as natural as breathing. The compliant scientific methodist must struggle to stifle what his cognitive nature has already made him believe. Doing so takes effort (and often time); so costs are necessarily levied.

Generic summations do not exhaust the class of non-universal generalizations. Normalic generalizations, of which statistical generalizations are a particular case, also figure prominently in ampliative reasoning. Normalic generalizations are generalizations about what is the case nearly always, or for the most part. There is a use of the word "normally" which is a synonym of

"usually", which our term "normalic" draws upon. Unlike generic generalizations, normalic generalizations embed quantifiers. This is not everyone's understanding of genericity. But in light of the fact that some claims of the form "*F*s *G*" are true and "Most *F*s are *G*" is false (Carlson and Pelletier, 1995), we think it the correct understanding. Genericizations lack a quantificational organization precisely where normalic generalizations have it essentially. It is an important structural difference, carrying interesting semantic consequences. Whereas "This *F* doesn't *G*" can be true without "*F*s *G*" ceasing to be true, it remains the case that "This *F* doesn't *G*" is a negative instance of "*F*s *G*", albeit a true one. Yet "This *F* doesn't *G*" doesn't come close to being a negative instance of "Nearly all *F*s *G*". How to fill in these semantic differences is still an open question in the logic of general propositions. Interesting and important though the question is, we shall not press it here. It suffices to note that

Proposition 20 (Variable generality)**.** *Thin-slicing carries no* intrinsic *tie to types of generalization.*

Accordingly, one may hastily generalize to Hempelian generalizations, generic generalizations and normalic generalizations. We have pointed out the advantages of genericizing over Hempelianizing. Like advantages attach to normalicizing rather than Hempelianizing. In each case, the truth of propositions in the form "This *F* doesn't *G*" needn't disturb the truth of the respectively generalization. This leaves the question as to what would differentially motivate generic and normalic thin-slicing. The answer, broadly speaking, hinges on the element of defectiveness. Negative instances of generic thin-slicing are in some or other way defective cases of the subject term. There is no such assumption to be made in the case of normalic thin-slicing.

Normalic thin-slicing is but one example of judgements of non-universal quantification. If we allow that "Nearly all" as a quantifier, "Hardly any" cannot be denied the same recognition. The difference between "Nearly all" and "Hardly any" mimics the difference between "*n*% of" and *m*% of", where *n* is quite large and *m* is quite small. So statistical projections also have the general character of non-universal quantification.

We see in these similarities and differences an important moral.

Proposition 21 (Low non-universality)**.** *"Fs are hardly ever G" is as much a case of thin-slicing as is "Fs are G" or "Fs are nearly always G" when drawn from a small (enough) sample.*

Thin-slicing is largely automatic. To a considerable extent, it is part of what goes on down below. Hasty generalization is also a belief-forming device; and, as we have seen, belief from the inside perspective manifests itself as knowledge. This would be an epistemic disaster if the hasty generalizations we actually are drawn to make were always or frequently mistaken. If so, we would

be massively mistaken in what we are induced to think that we know. What is so striking about hasty generalizations, as they are drawn in real life by beings like us, is that they are by and large right, or right enough to allow us to survive and prosper, to contribute to the replication of our cognitive devices in the human descendent class, and occasionally to build great civilizations. So we may say that

Proposition 22 (The naturalness of hasty generalization)**.** *The hasty generalizations actually drawn by practical agents are cheap, irresistible and typically accurate enough to* fulfil *our interests.*

We may hypothesize that the capacity for generally accurate generalizational haste is something that is hard-wired into beings like us, or that, in any event, it is so primitive a skill that it must have been part of the yield of our earliest learning. It doesn't matter. Once the human individual is past his early infancy, his life is saturated with generalizations that are both hasty and accurate, and, when not accurate, efficiently corrigible. It is tempting to speculate that it all springs from the mechanisms of flight and fight. Perhaps this is so. But, again, what matters for the logicians are not the causes of such haste, but the cognitive utilities of it.

For this unfolding apologia to be defensible, it must be the case that

Proposition 23 (Practicality and haste)**.** *The extent to which an agent is operating practically, is not by and large appropriate that his targets be such as to impose the standard of inductive strength.*

Let us pause to consider the view that we are trying to dispel.

1. Cognitive rationality is the system of thought prescribed by the deductive and inductive logic and decision theory.

2. Human beings are naturally so constituted that they think in ways that closely approximate to the canons prescribed by these systems.

3. Accordingly, a theory of rationality should provide an account of how the state of affairs stated by (2) came to mirror the norm expressed by (1).

Our position is that the norm embodied in (1) is no norm and the fact expressed by (2) is no fact.[20] If we want to be right in our rejection of the norm purported by (1), we must discourage the idea that beliefs sanctioned by the standard of inductive strength constitute a kind of global maximum. But, as we have already pointed out, there are reasons to doubt any such claim. Unlike

[20]Matthen shares our scepticism about (2). He is rather more equivocal about (1). See Matthen (2002, Section 6).

(classical) validity, which is wholly impervious to new information, inductive strength is a veritable sitting duck. We can see this in an especially dramatic way when *C* is a generalization and *E* is a sample. Like the universally quantified conditional construal of generality, the property of inductive strength is highly brittle. Let a given such argument be as inductively strong as may be. If the next bit of information is a counterexample *N* to *C*, the original argument remains inductively strong and the result of supplementing its premisses by addition of *N* is an inductive disaster. What this shows is that the inductive strength of the original argument was no reason to think well of it, whereas the catastrophe engineered by the present argument invests over-heavily in freedom from counterexample in inductive contexts. Thus the norm embodied in (1) can't be relied upon unless accompanied by reasonable assurances of the non-existence of counterexamples. But this asks more from ampliative reasoning than it can possibly be expected to provide.

It is instructive to compare ampliative reasoning in an individual's hands and in NASA's. NASA's targets are such that it must pay for its counterexamples with disasters. When an *N* comes along that topples a *C*, all bets are off until, with considerable elaboration, *C* is reframed so as to tolerate *N* or *N* is reformulated to take the pressure off *C*, or *C* is abandoned and hopes for a happier successor are launched. In actual practice, these accommodations are often very difficult and very expensive. Individuals by and large simply aren't up to these levels of disaster-management. Accordingly, individuals do not typically repose their ampliative burdens on so fickle a standard as inductive strength. Rather they show their fondness for genericity and the like, which in turn is an invitation to make do with small samples. This makes a nonsense of inductive strength, needless to say. But it gives the practical reasoner a form of ampliation that serves him well and that he can afford. For, again, he is not typically wrong in the generic claims he wrests from small samples with such haste; and when he is wrong, i.e., when a true negative instance *N* does present itself, he is not, just so, faced with the burden and the cost of repairing *C*. As we said, *C* is elastic; it can remain true in the face of true negative instances.

4.16 Risk aversion

Hasty generalization genericizes or quantifies from small samples. Doing so would clearly be defective if the samples in question were unrepresentative. In the literature on inductive logic, it is common to require of an agent that, before he generalizes from a sample, he check it for, or otherwise assure himself of, its representativeness. This is true but unhelpful. In generalizing from a small sample, a reasoner implies that the sample is representative. To make it a condition on such generalizations that they be grounded in the conviction that the sample is representative is to require him to withhold his generalization until he thinks that he has reason to think it correct.

What counts here is that thin-slicers—that is, all of us—are adept at discerning representative samples among the very small. We have already made the point that our facility with sample representativeness is linked to our facility with natural kinds. Doubtless this is so, but it doesn't amount to much of an explanation. Better that we explore the link with our danger-recognition capacities. Hard-wired or not, one of the most primitive and successful of an individual agent's endowments is the wherewithal for the timely recognition of danger even in the face of utterly scant evidence of it. The attendant protocols of risk aversion are concomitantly *conservative.* They risk the effort of unnecessary evasions for the advantage of securing against the greater liabilities that attach to the contrary. The flight-fight mechanisms of beings like us are activated by factors of apprehensiveness; fear is the third 'f' in this trio. They are mechanisms that embed the fundamental structure of thin-slicing.

The fear factor is crucially important. When an individual runs from the unknown creature with large fangs, it is not at all necessary to attribute to him the tacit belief that such creatures are lethal biters but rather the anxiety that they *might* be. Risk aversion turns on epistemic estimates of comparatively low yield; not on the conviction that *F*s *G* but on the worry that *F*s might *G*. Behaviour is risk-averse in this conservative way precisely when it grounds non-trivial action on so slight and tentative an appreciation of what is the case. We may see thin-slicing as an adaptation of conservative rise-averse behaviour, in which the element of fear is replaced by that of belief and the estimate of mere possibility is upgraded somewhat. Even so, the basic structure is retained. When on the strength of a small sample one reasons that *F*s *G* or (most do or few do), one is tendering the projection with a requisite tentativeness. But if this is so, thin-slicing cannot be judged by the standard of inductive strength.

4.17 Probabilistic reasoning

Given that an argument is inductively strong to the extent that its conclusion is made more likely by the evidence cited in the premisses, a number of additional assumptions are the life's blood of mainstream inductive logic.

1. Likelihood is *probability.*

2. The relation of greater (or less) likelihood relative to a body of evidence is the relation of *conditional probability.*

3. The concepts of probability and conditional probability are accurately described by the theorems of the *probability calculus.*

4. Any set of premises that increases the conditional probability of a proposition also confers some positive degree of *confirmation* on it.

We have tried to make plain that the inductive strength standard is neither appropriate nor required for a practical agent's cognitive targets by and large. What would count against this claim? Here is a point that might give us pause. Everyone agrees that practical agents have an impressive command of probabilistic reasoning. Suppose it turned out that the present assumptions are true, and that actual probabilistic reasoning comported with them. If these things were so, our real-life probabilistic reasoning would satisfy conditions under which probabilistic success would indeed hit the standard of inductive strength. Clearly, we must say something about probabilistic reasoning.

If the behaviour of individual agents is anything to go one, then the standard accounts of inductive inference constitute significant distortion of the actual record. Can the same be said for the linked issue of probabilistic reasoning in the here-and-now? James Franklin sees in probability an interesting parallel with continuity and perspective (Franklin, 2001). All three of these things took a long time before yielding to mathematical formulation, and, before that happened, judgements of them tended to be unconscious and mistaken. We have a somewhat different version of this story. Sometimes a conceptually inchoate idea is cleaned up by a subsequent explication of it. Sometimes these clarifications are achieved by modelling the target notion mathematically. Sometimes the clarification could not have been achieved save for the mathematics. We may suppose that something like this proved to be the case with perspective and continuity. To the extent that this is so, anything we used to think of these things which didn't make its way into the mathematical model could be considered inessential if not just mistaken. It is interesting to reflect on how well this line of thought fits the case of probability.

In raising the matter, we are calling attention to two questions. (1) What was probability like before Pascal? (2) How do we now find it to be? Concerning the first of this pair of questions, We think that we may suppose that, in their judgements under conditions of uncertainty, people routinely smudged such distinctions as may have obtained between and among 'it is probable that', 'it is plausible that', and 'it is possible that'. If we run a strict version of that line over this trio, then not making it into the calculus of Probability leaves all that is left of these blurred idioms in a probabilistically defective state. There is a sense in which this is not the wrong thing to conclude, but it is a trivial one. For if what we sometimes intend by 'probability' fails to find a safe harbour in the probability calculus, then it is not a fact about probability that the probability calculus honours. But unlike what may have been the case with perspective and continuity, we must take care not to say without further ado that those inferences that don't make the Pascalian cut are mistakes of reason or even mistakes of probabilistic reason. In this we cast our lot with Cohen (1982) and Toulmin (1953) albeit for somewhat different reasons. With Cohen we agree that some of the Kahneman and Tversky (1974) experimental

results which show their subjects to have been bad Pascalians do so only if they had undertaken to be good Pascalians. The alternative, of course, is that, even though they were invited to be Pascalians and primed to make a workmanlike job of it, their sole mistake is that they slid unawares into a non-Pascalian disposition toward reasoning under conditions of uncertainty. Certainly had they been drawn to the task of compounding *plausibilities,* it is far from clear that the Kahneman-Tversky results show their efforts in a bad light.

We side with Toulmin in saying that not all judgements of probability, even when made by working scientists, express or attempt to express the concept of aleatory probability or to comport with its theorems. A similar moral can be drawn from the sheer semantic sprawl of the idioms of possibility.

Let us take it that, unlike perspective and continuity, idioms of probability (or probability/plausibility/possibility) that don't cut the Pascalian mustard leave residues of philosophically interesting usage. If this were so, there might well be philosophically important issues, the successful handling of which requires the wherewithal of this conceptual residue. Again, standard answers to Kahneman-Tversky questions don't cut the mustard of aleatory probability, but they do comport with conditions on plausible reasoning. What, then, are we to say? That these bright, well-educated subjects are Pascalian misfits or that they are more comfortably at home (though unconsciously) with a plausibility construal of their proferred tasks? If we say the second, we take on an onus we might be unable to discharge, or anyhow discharge at will. It is the task of certifying the conditions under which these non-Pascalian manoeuvres are well-justified. In lots of cases, we won't have much of a clue as to how to achieve these elucidations. Small wonder, then, that what we call the Can Do Principle beckons so attractively. This is the principle that bids the theorist who is trying to solve a problem *P* to stick with what he knows and to make a real effort to adapt what he knows to the requirements of *P*. One of the great attractions of Pascalian probability is that we know how to axiomatize it. Can Do is right to say that it would be advantageous if we could somehow bend the probability calculus to the task to hand. But sometimes, the connection just can't be made.

Bas van Fraassen is spot on in pointing out that there "has been a sort of subjective probability slum in philosophy, and its inhabitants, me included, have not convinced many other philosophers that what happens there is anything more than technical self-indulgence" (van Fraassen, 2005). This calls to mind our *Make* Do Principle, which is the degenerate case of Can Do. Make Do is just Can Do in circumstances in which the fit with *P* cannot be achieved satisfactorily. If *P* is the problem of avoiding "the naïveté and oversimplification inherent in much of traditional epistemology" (van Fraassen, 2005), then a decision to deploy the theory of probability by brute force would be a case of Make Do. It would capture the mood of the tasker who, not knowing what to

do about P, settles for he knows how to do about Q, and wholly ignores that it is all beside the point.

4.18 The link to abduction

If what we have been saying about thin-slicing is correct, hasty generalization bears a significant resemblance to abductive reasoning. Abductive reasoning is a response to an ignorance-problem. One has an ignorance-problem when one has a cognitive target that cannot be attained on the basis of what one currently knows. Ignorance problems trigger one or other of three responses. In the one case, one overcomes one's ignorance by attaining some additional knowledge. In the second instance, one yields to one's ignorance (at least for the time being). In the third instance, one abduces. The general form of an abductive inference can be set out as follows, putting T for the agent's target, K for his (or its) knowledge-base, K^* for an accessible successor-base of K,[21] R as the attainment relation relative to T, H as the agent's hypothesis; $K(H)$ as K's adaptation of H, and R^{pres} as the relation of presumptive attainment relative to T:

1.	$\neg R(K, T)$	[fact]
2.	$\neg R(K^*, T)$	[fact]
3.	$R^{pres}(K(H), T)$	[fact]
4.	Therefore, $C(H)$	[conclusion]
5.	Therefore, H^c	[conclusion]

What the schema tells us is this: T cannot be attained on the basis of Q. Neither can it be attained on the basis of any successor K^* of K that the agent knows then and there how to construct. H is an hypothesis such that when reconciled to K produces $K(H)$. H is such that if it were true, then $K(H)$ *would* attain T. But since H is only hypothesized, its truth is not assured. Accordingly we say that $K(H)$ *presumptively* attains T. That is, having *hypothesized* that H, the agent *presumes* that his target is now attained. But since presumptive attainment isn't attainment, the agent's abduction must be seen as preserving the ignorance that gave rise to his (or its) ignorance-problem in the first place. Accordingly, abduction is not a *solution* of an ignorance problem, but rather a *response* to it, in which the agent settles for presumptive attainment rather than attainment. $C(H)$ expresses the conclusion that it follows from the facts of the schema that H is a worthy object of conjecture. H^c denotes the decision

[21] K^* is an accessible successor to K to the degree that an agent has the know-how to construct it in a timely way; i.e., in ways that are of service in the attainment of targets aimed at K. For example if I want to know how to spell "accommodate", and have forgotten, then my target can't be hit on the basis of K what I now know. But I might go to my study and consult the dictionary. This is K^*. It solves a problem originally aimed at K.

to release H for further promissory work in the domain of enquiry in which the original ignorance-problem arose. The superscript is a label. It reminds us that H has been let loose on sufferance. (For an exhaustive discussion of abduction, see Gabbay and Woods, 2005.)

Abductions are a response to ignorance-problems intermediate between solving them and being defeated by them. Like the latter, successful abductions do not solve the ignorance-problems that give them rise. Like the former, abductions authorize (albeit defeasibly) subsequent actions that the agent may well have preferred to have seen grounded in a solution to his problem, rather than in an ignorance-preserving accommodation of it. Even so, abductions do license inferences, on which subsequent actions might reasonably be taken (albeit defeasibly, both times).

Thin-slicing resembles abduction in certain quite clear ways. Just as the relation between $K(H)$ and T is *presumptive* attainment for the abducer, so, for the thin-slicer, is the inference from his small sample and his generalization *presumptive.* Just as the abducer's inference of $C(H)$ itself only a *plausible inference,* the thin-slicer's instantiation of his generalization is a *default.*

What is less clear is whether it is invariably the case that whenever a thin-slicer slices thinly he has (however tacitly) set himself an abductive target. Certainly we may assume that when faced with a small sample, no hasty generalizer will take the view that this constitutes a knowledge-base that attains the cognitive target (if that's what it is) of *knowing* that the generalization is true. But it is another thing entirely as to whether we might also assume that in reaching his more qualified inference—presumptive generalization, as we might call it,—there is an hypothesis H which, when added to the sample, would indeed attain the generalization unqualifiedly. Of course, there is such an H. It says that the sample is *representative.* But it won't quite do for abduction, since the proposition that the sample of Fs that G is representative just is the generalization in question.

No doubt, these and other questions could be explored to advantage. But, unless we are mistaken, we have already seen enough of the similarity between thin-slicing and abduction to be able to emphasize what is essential to the making of hasty generalizations. They are made *presumptively,* and the instantiations they sanction are *defaults.* This gives us what we want.

Proposition 24 (Confirming our thesis)**.** *Individual thin-slicers characteristically do not take on the standard of inductive strength. Given that an inductive fallacy is one that fails the standard of inductive strength, the* GOE *fallacy of hasty generalization is characteristically not a fallacy committed by beings like us.*

What is more,

Proposition 25 (The virtue of haste)**.** *Given the generally good track record of individual thin-slicers and the considerable economies that thin-slicing achieves, the practice of hasty generalization possesses, for beings like us, the cognitive virtue of producing large stores of default propositions on which to ground, with due regard for the attendant risks, the appropriate actions. In other words, thin-slicing is a natural discouragement of paralyzing indecision.*

As we have remarked, the slenderness of our own sample might well leave the reasonable reader unmoved to accept our ambitious claim for all 18 members of *GOE*. Certainly, we will not be so brazen as to suggest that sample produced by thin-slicing is representative of all of *GOE*. Let us say it again. It is a small sample. Perhaps we would have been better advised to entitle our paper "Thin-slicing as a cognitive virtue". Even so, we do think that some quite general lessons can be drawn from our examinations of this sample. One is that a piece of reasoning is a fallacy only in relation to what the agent has in mind to achieve cognitively. So, at a minimum, before we can rightly accuse an individual agent of committing a *GOE*-fallacy, we must have independent reasons for supposing that his target T carries standards S that his reasoning violates. In light of the forgoing discussion, we take it as given that it is often far from obvious that such Ts and Ss are actually in play in the cognitive lives of beings like us.

We keep saying "beings like us". This is because it matters. Beings like us are individual agents. Individual agents tend to set for themselves moderate targets. Moderate targets are those that can be attained (or as we may now say, presumptively attained) by the deployment of scant-resources i.e., scant in comparison to what NASA and MI5 command. Agents whose resource-draw is greatly larger than ours certainly set themselves tougher targets governed by higher standards. We don't doubt for a minute that when NASA was in process of generalizing about O-ring integrity, it was clearly targeted on scientific certainty and clearly pledged to the standard of inductive strength. In such circumstances, thin-slicing would have been cognitively defective; worse, it would have been an ethical catastrophe. There is also a moral to be drawn from this.

Proposition 26 (Vindication of the tradition)**.** *On the traditional view, a fallacy is an inapparent error. Leaving aside the general point that all errors are inapparent, we see that hasty generalization conforms to the traditional view. For it* is *an error (when committed by NASA) and it* looks not to be *an error, because it is not an error (when committed by beings like us).*

Finally, we should make brief mention of the Principle of Charity.[22] The Principle of Charity bids us not to interpret our interlocutors in ways that convict them of error or irrationality; more carefully, we are not so to interpret them except in default of strong indications to the contrary. The Principle of Charity is itself hardly free from controversy, and we have no wish to rush to judgement on its behalf (see Woods, 2004, Chapter 14). Suffice it to say that if, when done by us, thin-slicing is indeed a fallacy, then beings like us are *massive inductive misfits.* There is, apart from the soundness of the Charity Principle, a further reason to doubt it. Suppose that we were indeed massive inductive misfits. It would hardly matter. For we get things more right than wrong. We survive, we prosper, and occasionally we build great civilizations. What this would tell us, given the present assumption, is that it is *not irrational* to be massive inductive misfits.

Acknowledgments

Our work is supported by the Engineering and Physical Sciences Research Council of the United Kingdom, the Social Sciences and Humanities Research Council of Canada, the Dean of Arts, University of British Columbia, the Dean of Arts and Science, University of Lethbridge, and the Head of Computer Science at King's College. We are deeply grateful for this assistance. We also extend our gratitude to Carol Woods for invaluable technical support. It is a pleasure to record our debt to the many persons with whom conversations and correspondence have helped us materially in forming our views about fallacies and the logic of cognitive systems: Atocha Aliseda, Peter Alward, Peter Bruza, Balasishnan Chandrasekaran, Artur Garcez, Hans Hansen, Jaakko Hintikka, David Hitchcock, Scott Jacobs, Erik Krabbe, Theo Kuipers, Henrike Jansen, Luis Lamb, Dom Lopes, Peter McBurney, Mohan Matthen, Lorenzo Magnani, Sami Paavola, Kent Peacock, Jeanne Peijenburg, Ahti-Veikko Pietarinen, Patrick Rysiew, Matti Sintonen, Patrick Suppes, Stephen Toulmin, Bas van Fraassen, Johan van Benthem, Paul Viminitz, Mark Weinstein, Joseph Wenzel.

References

Alchourron, C. A., Gärdenfors, P. G., and Makinson, D. (1985). On the logic of theory change; partial meet, contraction and revision functions. *The Journal of Symbolic Logic*, 50:510–530.

Ambady, N. and Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical atractiveness. *Journal of Personality and Social Psychology*, 64:431–441.

[22]Charity is way of distinguishing between rival analytical hypotheses, which is Quine's word for a translation manual for an alien's linguistic behaviour (Quine, 1970, Chapter 2) and concerning which Davidson holds that "[c]harity is forced upon us; whether we like it or not, if we want to understand others, we must count them right in most matters" (Davidson, 1984, xviii). Scriven writes to the same effect that the "Principle of Charity requires that we make the best, rather than the worst possible interpretation. . . ." (Scriven, 1976, 71).

Barth, E. M. and Krabbe, E. C. W. (1982). *From Axiom to Dialogue*. de Gruyter, Berlin, New York.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. MIT, Cambridge, MA.

Bermúdez, J. L. (2004). *Thinking Without Words*. Oxford University Press, Oxford.

Black, M. (1946). *Critical Thinking*. Prentice-Hall, New York.

Bruza, P. D., Song, D., and McArthur, R. M. (2004). Abduction in semantic space: Towards a logic of discovery. *Logic Journal of the IGPL*, 12(2):97–109.

Bruza, P. D., Cole, R. J., Song, D., and Abdul Bari, K. (2006) Towards operational abduction from a cognitive perspective. *Logic Journal of the IGPL*, 14(2):161–177.

Carlson, G. N. and Pelletier, F. J., editors (1995). *The Generic Book*. Chicago University Press, Chicago, IL.

Carney, J. D. and Scheer, R. K. (1980). *Fundamentals of Logic*. Macmillan, New York, 3rd edition.

Carrere, S. and Gottman, J. (1999). Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family Process*, 38(3):293–301.

Chellas, B. (1980). *Modal Logic: An Introduction*. Cambridge University Press, Cambridge.

Churchland, P. (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. MIT, Cambridge MA.

Churchland, P. (1995). *The Engine of Reason, the Seat of the Soul*. MIT, Cambridge, MA.

Cohen, J. (1982). Are people programmed to commit fallacies: Further thoughts about the interpretation of experimental data and probability judgement. *Journal of Theory and Social Behavior*, 12:251–274.

Copi, I. M. (1986). *Introduction to Logic*. Macmillan, New York, 7th edition.

Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford University Press, Oxford, New York.

d'Avila Garcez, A. S. and Lamb, L. C. (2004). Reasoning about time and knowledge in neural-symbolic learning systems. In Thrum, S. and Schoelkopk, B., editors, *Proceedings of the NIPS 2003 Conference*, volume 16 of *Advances in Neural Information Processing Systems*, Vancouver, BC, MIT, Cambridge, MA.

d'Avila Garcez, A. S., Broda, K., and Gabbay, D. M. (2002). *Neural-Symbolic Learning Systems: Foundations and Applications*. Springer, Berlin.

Franklin, J. (2001). *The Science of Conjecture: Evidence and Probability Before Pascal*. The Johns Hopkins University Press, Baltimore, MD.

Freeman, J. B. (1991). *Dialectics and the Microstructure of Argument*. Foris, Dordrecht.

Frege, G. (1879). *Begriffschrift, a Formal Language, Modeled upon That of Arithmetic, for Pure Thought*. Harvard University Press, Cambridge, MA.

Gabbay, D. M. and Woods, J. (2001a). More on non-cooperation in dialogue logic. *Logic Journal of the IGPL*, 9:321–339.

Gabbay, D. M. and Woods, J. (2001b). The new logic. *Logic Journal of IGPL*, 9:157–190.

Gabbay, D. M. and Woods, J. (2001c). Non-cooperation in dialogue logic. *Synthese*, 127:161–186.

Gabbay, D. M. and Woods, J. (2003a). *Agenda Relevance: A Study in Formal Pragmatics*, volume 1 of *A Practical Logic of Cognitive Systems*. North Holland, Amsterdam.

Gabbay, D. M. and Woods, J. (2003b). Normative models of rationality: The theoretical disutility of some approaches. *Logic Journal of IGPL*, 11:597–613.

Gabbay, D. M. and Woods, J. (2005). *The Reach of Abduction: Insight and Trial*, volume 2 of A Practical Logic of Cognitive Systems. Elsevier, Amsterdam.

Gabbay, D. M. and Woods, J. (2009). Errors of logic. Forthcoming.

Gabbay, D. M. (1976). *Investigations in Modal and Tense Logics with Applications*. Reidel, Dordrecht, Boston, MA.

Gabbay, D. M., Hodkinson, I., and Reynolds, M. (1994). *Temporal Logic: Mathematical Foundation and Computational Aspects*, volume 1. Oxford University Press, Oxford.

Gabbay, D. M., Rodrigues, O., and Woods, J. (2002). Belief contraction, antiformulae and resourse-overdraft: Part I: Deletion in resources bounded logics. *The Logic Journal of the IGPL*, 10:601–652.

Gabbay, D. M., Pigozzi, G., and Woods, J. (2004a). Controlled revision: A preliminary account. *The Logic Journal of the IGPL*, 13:5–27.

Gabbay, D. M., Rodrigues, O., and Woods, J. (2004b). Belief contraction, antiformulae and resourse-overdraft: Part II: Deletion in resources unbounded logics. In Rahman, S., Symons, J., Gabbay, D. M., and van Bendegem, J. P., editors, *Logic Epistemology and the Unity of Science*, pages 291–326. Kluwer, Dordrecht, Boston, MA.

Gigerenzer, G. and Selten, R. (2001). Rethinking rationality. In *Bounded Rationality: The Adaptive Toolbox*, pages 1–12. MIT, Cambridge, MA.

Gochet, P. (2002). The dynamic turn in twentieth century logic. *Synthese*, 130:175–184.

Gochet, P. and Gribomont, P. (2005). Epistemic logic. In Gabbay, D. M. and Woods, J., editors, *Handbook of the History of Logic*, vol 7, pages 99–195. Elsevier, Oxford.

Grootendorst, R. (1987). Some fallacies about fallacies. In van Eemeren, F. H., Grootendorst, R., Blair, J. A., and Willard, C. A., editors, *Argumentation Across the Lines of Discipline*, pages 331–342. Foris, Dordrecht, Providence.

Guarini, M. (2001). A defence of connectionism against the "SYNTACTIC" argument. *Synthese*, 128:287–317.

Hamblin, C. L. (1970). *Fallacies*. Methuen, London.

Hansen, H. V. and Pinto, R. C., editors (1995). *Fallacies: Classical and Contemporary Readings*. Pennsylvania State University Press, University Park, PA.

Harel, D. (1979). *First-Order Dynamic Logic*. Springer, Berlin.

Harman, G. (1986). *Change in View: Principles of Reasoning*. MIT, Cambridge, MA.

Hilpinen, R., editor (1981). *Deontic Logic: Introductory and Systematic Readings*. Reidel, Dordrecht.

Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press, Ithaca, NY.

Hintikka, J. (1981). *Modern Logic—A Survey*. Reidel, Boston, MA.

Hintikka, J. (1987). The fallacy of fallacies. *Argumentation*, 1:211–238.

Hintikka, J. (1997). What was Aristotle doing in his early logic, anyway? A reply to Woods and Hansen. *Synthese*, 113:241–249.

Hintikka, J. and Sandu, G. (1997). Game-theoretical semantics. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 361–410. Elsevier, Amsterdam.

Horgan, T. and Tienson, J. (1999a). Authors' replies. *Acta Analytica*, 22:275–287.

Horgan, T. and Tienson, J. (1999b). Short précis of connectionism and the philosophy of psychology. *Acta Analytica*, 22:9–21.

Irvine, A. D. (1989). Epistemic logicism and Russell's regressive method. *Philosophical Studies*, 55:303–327.

Johnson, O. (1967). Begging the question. *Dialogue*, 6:135–160.

Johnson, R. H. (1996). *The Rise of Informal Logic*. Vale, Newport News, VA.

Johnson, R. H. (2000). *Manifest Rationality: A Pragmatic Theory of Argument*. Lawrence Erlbaum Associates, London.

Kahneman, D. and Tversky, A. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185:1124–1131.

Kowalski, R. A. (1979). *Logic for Problem Solving*. Elsevier, New York.

Kripke, S. A. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica*, 83–94.

Lenzen, W. (1978). Recent work in epistemic logic. *Acta Philosophica Fennica*, 30:1–219.

MacKenzie, J. (1990). Four dialogue systems. *Studia Logica*, XLIX:567–583.

Magnani, L. (2001). *Abduction, Reason and Science: Processes of Discovery and Explanation*. Kluwer, Plenum, New York.

Massey, G. J. (1981). The fallacy behind fallacies. *Midwest Studies in Philosophy*, 6:489–500.

Matthen, M. (2002). Human rationality and the unique origin constraint. In Ariew, A., Cummins, R., and Perlman, M., editors, *Functions: New Readings in the Philosophy of Psychology and Biology*, pages 341–372. Oxford University Press, Oxford.

McCarthy, J. (1980). Circumscription—A form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39.

Moore, R. (1985). Semantical considerations on non-monotonic logics. *Artificial Intelligence*, 25:75–94.

Peirce, C. S. (1992). *Reasoning and the Logic of Things: The Cambridge Conference Lectures of 1898*. Ketner, K. L., editor, introduction by Kenneth Laine Ketner and Hilary Putnam. Harvard University Press, Cambridge, MA.

Pereira, L. M. (2002). Philosophical incidence of logic programming. In Gabbay, D. M., Johnson, R. H., Ohlbach, H. J., and Woods, J., editors, *Handbook of the Logic of Argument and Inference: The Turn Towards the Practical*, volume 1, pages 421–444. North Holland, Amsterdam.

Powers, L. (1995). Equivocation. In Hansen, H. V. and Pinto, R., editors, *Fallacies: Classical and Contemporary Readings*, pages 287–301. Pennsylvania State University Press, University Park, PA.

Prior, A. N. (1967). *Past Present and Future*. Oxford University Press, Oxford.

Quine, W. V. O. (1970). *Philosophy of Logic*. Prentice Hall, Englewood Cliffs, NJ.

Raiffa, H. (1968). *Decision Analysis*. Addison Wesley, Reading, MA.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 12:81–132.

Rensink, R. (2000). Visual sensing without seeing. *Psychological Science*, 15:27–32.

Sandewall, E. (1972). *An Approach to the Frame Problem and Its Implementation*. Edinburgh University Press, Edinburgh.

Schipper, E. W. and Schuh, E. (1959). *A First Course in Modern Logic*. Henry Holt, New York.

Schlechta, K. (2004). *Coherent Systems*. Elsevier, Amsterdam.

Scriven, M. (1976). *Reasoning*. McGraw-Hill, New York.

Sorensen, R. A. (1988). *Blindspots*. Clarendon, Oxford.

Tarski, A. (1956). The concept of truth in formalized languages. In *Logic Semantics, Metamathematics*, pages 152–278. Translated by J. H. Woodger. Clarendon, Oxford.

Toulmin, S. (1953). *The Philosophy of Science: An Introduction*. The Hutchinson University Library, London.

van Benthem, J. (1983). *The Logic of Time*. Reidel, Dordrecht.

van Benthem, J. (1996). *Exploring Logical Dynamics*. CSLI, Stanford.

van Fraassen, B. C. (2005). The day of the dolphins: Puzzling over epistemic partnership. In Peacock, K. A. and Irvine, A. D., editors, *Mistakes of Reason: Essays in Honour of John Woods*, pages 111–133. University of Toronto, Toronto.

von Wright, G. H. (1951). *An Essay in Modal Logic*. North Holland, Amsterdam.

Walton, D. (1995). *A Pragmatic Theory of Fallacy*. University of Alabama Press, Tuscaloosa, AL.

Walton, D. and Krabbe, E. C. W. (1995). *Commitment in Dialogue*. SUNY, Albany, NY.

Wheeler, M. (2001). Two threats to representation. *Synthese*, 129:211–231.

Williamson, J. (2002). Probability logic. In Gabbay, D. M., Johnson, R., Ohlbach, H. J., and Woods, J., editors, *Handbook of the Logic of Argument and Inference: The Turn Towards the Practical*, pages 397–424. North Holland, Amsterdam.

Woods, J. (2003). *Paradox and Paraconsistency: Conflict Resolution in the Abstract Sciences*. Cambridge University Press, Cambridge.

Woods, J. (2004). *The Death of Argument: Fallacies in Agent-Based Reasoning*. Kluwer, Dordrecht, Boston.

Woods, J. and Hansen, H. V. (1997). Hintikka on Aristotle's fallacies. *Synthese*, 113:217–239.

Woods, J. and Walton, D. (1989). *Fallacies: Selected Papers 1972–1982*. Foris–de Gruyter, Berlin, New York.

Woods, J., Irvine, A., and Walton, D. (2004). *Argument: Critical Thinking, Logic and the Fallacies*. Prentice Hall, Toronto, 2nd edition.