MICHAEL A. BISHOP

THE AUTONOMY OF SOCIAL EPISTEMOLOGY[1]

Social epistemology is autonomous: When applied to the same evidential situations, the principles of social rationality and the principles of individual rationality sometimes recommend inconsistent beliefs. If we stipulate that reasoning rationally from justified beliefs to a true belief is normally sufficient for knowledge, the autonomy thesis implies that some knowledge is essentially social. When the principles of social and individual rationality are applied to justified evidence and recommend inconsistent beliefs and the belief endorsed by social rationality is true, then that true belief would be an instance of social knowledge but not individual knowledge.

This paper is organized into six sections. Section 1 describes a widely held conception of individual epistemology and the individualistic notion of rationality that drops out of it. A principle of rationality that has helped to shape judgments made within the contexts of social institutions is introduced in section 2. Sections 3-5 argue that this principle of social rationality is practically inconsistent with individual rationality. The final section concludes with a brief discussion of some objections and implications.

1. Epistemological Individualism

Epistemological individualism consists of two widely accepted theses. (1) *Methodological individualism* holds that knowledge of the correct epistemic norms is somehow available to individuals and we can discover that knowledge via an appropriate process of self-exploration (Goldman and Pust 1994, Bealer 1987, BonJour 1998). (2) *Metaphysical individualism* holds that only individual cognizers can have knowledge (for a holdout, see Schmidtt 1994).

Methodological individualism is presupposed in the method of contemporary analytic epistemology, where successful theories are supposed to capture our epistemic intuitions about well-developed cases. Epistemic intuitions are usually taken to be our non-discursive, though perhaps considered, judgments about the epistemic properties of some cognitive item (Cohen 1981, Bealer 1987, BonJour 1998, Pust 2000). Paradigm examples of epistemic intuitions are our judgments that subjects in Gettier cases do not have knowledge. It is useful to think about methodological individualism in terms of a performance-competence distinction. We tacitly possess, or in some way have access to, a reasoning competence – a set of principles about how we ought to reason. These principles define individual rationality. As a result, "ordinary human reasoning… cannot be held to be faultily programmed: it sets its own standards" (Cohen 1981, 317). How, then, are we to explain the fact that people sometimes reason poorly? Reasoning involves bringing to bear not only our reasoning competence but also our background beliefs and various ancillary cognitive systems, such as memory, attention and perception. Performance errors occur when we reason poorly as a result of a failure of our background beliefs or of our ancillary systems (e.g., see Cohen 1981, 323-326; Goldman and Pust 1994, 182-183).[2] The individualist view under consideration takes all reasoning errors to be performance errors. There can be no systematic flaws in our reasoning competence, since that competence defines individual rationality.[3]

Metaphysical individualism, which holds that only individual persons can have knowledge, naturally coheres with a consequentialist conception of social epistemology: the epistemic quality of a social practice is a function solely of its epistemic consequences *for individuals* (Solomon 2000). On this view, social epistemology is a purely applied field. As with many applied fields, it can be excruciatingly

https://doi.org/10.3366/epi.2005.2.1.65 Published online by Cambridge University Press

difficult, involving novel and elaborate conceptual and mathematical machinery. But theoretically, it is completely derivative on the epistemology of individual knowers. The case for autonomy works equally well if we plug any number of reasonable epistemic desiderata into the individualist's consequentialist scheme. The reliabilist approach is popular and has generated some interesting results (e.g., Kitcher 1990, Goldman 1999). I will therefore assume that social epistemology "is concerned with the truth-getting impact of different patterns and arrangements of social intercourse" (Goldman 1986, 5). Social rationality involves arranging social practices so as to maximize (significant) true belief and minimize (harmful) false belief. This is a very conservative conception of social epistemology. It ignores many challenging and important views about the nature of social epistemology. But it would hardly be interesting to show that social epistemology is autonomous if I began by assuming (say) that knowledge is socially constructed or essentially distributed across social groups. If I can show that social epistemology is autonomous *given these individualist-friendly assumptions*, then it is autonomous, period.

2. The Flat Maximum Principle

For over 50 years, psychologists have developed rules for making judgments about matters of practical importance. These Statistical Prediction Rules (SPRs) outperform human experts in making some important judgments. For example, SPRs are more reliable than forensic psychologists in predicting the likelihood of violence and expert criminologists in predicting criminal recidivism (Faust and Ziskin 1988; Carroll et al., 1988); they outperform clinicans at correctly diagnosing progressive brain dysfunction on the basis of intellectual tests, and predicting the presence, location and cause of brain damage (Leli and Filskov 1984; Wedding 1983). Paul Meehl's evaluation of this literature is apt: "There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing [scores of] investigations, predicting everything from the outcomes of football games to the diagnosis of liver disease and when you can hardly come up with a half dozen studies showing even a weak tendency in favor of the clinician, it is time to draw a practical conclusion" (1986, 372-3).

Consider a common SPR, a linear model for making a judgment about a target property, P, on the basis of predictor cues c_1–c_4:

$$P = w_1c_1 + w_2c_2 + w_3c_3 + w_4c_4$$

where c_n is the value for the n^{th} predictor cue, and w_n is the weight assigned to the n^{th} cue. Raw values are not plugged into the variables. They are first normalized (they are z-scored) and then positively correlated with the target property. For example, a linear model that predicts the quality of the vintage for a red Bordeaux wine might reflect the fact that Bordeaux grapes thrive best in hot dry summers (Ashenfelter, et al. 1995). So if the average rainfall in a given summer is one standard deviation above (below) the mean, the value for the rainfall cue might be −1 (+1).

Given predictor cues that are normalized and positively correlated with the target property, we can distinguish three different kinds of linear model according to how their weights are chosen (Dawes 1982).

1. In a *proper model*, weights are chosen so as to best fit the data. Imagine a graph in which the predictor cues are plotted against the target property. The proper linear model draws a straight line that best fits those points.
2. In a *unit model*, all the weights are equal to 1. Each cue has an equal "say" in the final prediction. This is an *improper linear model* because its weights are not chosen so as to best fit the data.
3. In a *random model*, the weights are assigned randomly (but between 0 and 1).

A proper model will have many correlate improper models. Many of these will differ from the proper model only in terms of their weights; but some will drop those cues used by the proper model that are weakly correlated with the target.

Most people are not particularly surprised that proper linear models outperform human experts. Such models make judgments that best fit the available evidence – something experts cannot

 https://doi.org/10.3366/epd.2005.2.1.65 Published online by Cambridge University Press

be expected to do on the fly. Even so, the reaction of wine-tasting experts to successful proper linear models has been "somewhere between violent and hysterical" (Passell 1990). Most people are surprised, however, at the accuracy of improper linear models. In a classic essay, Dawes and Corrigan (1974) found that *for problems of social prediction, random and unit models were about as reliable as proper models, and therefore more reliable than human experts*. Dawes reports that this result "when published engendered two responses. First, many people didn't believe them – until they tested out random and unit models on their own data sets. Then, other people showed that the results were trivial… I concur with those proclaiming the results trivial, but not realizing their triviality at the time, I luckily produced a 'citation classic'…" (1988, 209, n. 17). The result is trivial because of a finding in statistics called *the flat maximum principle* (Einhorn and Hogarth 1975).

The flat maximum principle (FMP) says that for certain sorts of problem, the weights (between 0 and 1) assigned to the predictor cues of a linear model don't affect the model's long term accuracy as long as:

1. The judgment problem is difficult. What is being predicted is messy – the data points are spread out enough so that no linear model will be especially reliable.
2. The evidential cues are reasonably predictive.
3. The evidential cues are somewhat redundant. As a result, adding extra cues to a successful model might add very little (if any) accuracy.

Problems of social judgment tend to share these features. So for such problems, the FMP implies that improper models will be about as reliable as proper models.

It is easy to underestimate the counterintuitive nature of the FMP. To fully understand its oddness, consider figure 1, which compares a Proper Model (PM) and an Improper Model (IM) applied to a binary (yes/no) problem. The FMP implies that IM and PM will be about equally reliable overall. *But it doesn't follow that they will always agree*. In fact, when proper and improper models are applied to the same set of problems, it is typical for one to be more reliable than the other. This is not inconsistent with the FMP for two reasons. First, two models might be equally reliable in the long run but not equally reliable on a short run of problems. Second, and less obviously, the FMP does not imply that a particular proper linear model will be exactly as reliable in the long term as all of its correlate improper models. The proper model and its correlate improper models will typically have a range of different long term reliability scores. What the FMP implies is that we should not expect proper models to be significantly more reliable *on average* than their correlate improper models.

	1	2	3	4	
	55%	*15%*	*15%*	*15%*	% Correct
Proper model (PM)	T	F	T	F	*70%*
Improper model (IM)	T	F	F	T	*70%*

Figure 1: Comparing a Proper Model (PM) and an Improper Model (IM)

Figure 1 depicts a hypothetical example in which a PM and IM agree *70%* of the time (columns 1 and 2). They're both right *55%* of the time (column 1) and both wrong *15%* of the time (column 2). When they disagree (columns 3 and 4), neither is more accurate than the other. (Again, this need not be the case.)

The FMP has primarily been used to help social institutions make more reliable judgments. For example, improper models are now standardly used by financial institutions in making loan decisions and setting credit limits (Stillwell et. al. 1983). The FMP is a principle of social rationality because it plays an essential role in arranging social practices that maximize true belief and minimize false belief. Any plausible social epistemology must make room for it.

3. Selective Defection Studies

The FMP is practically inconsistent with the principles of individual rationality: As a matter of fact, there are many cases in which the FMP and our reasoning competence, in the form of our epistemic intuitions, recommend inconsistent beliefs; further, the epistemic intuitions that clash

https://doi.org/10.3366/epi.2005.2.1.65 Published online by Cambridge University Press

with social rationality are not the result of performance errors (i.e., ignorance, false beliefs, or poor reasoning). Many readers will object because they believe it is irrational for us not to accept and reason in accordance with the FMP. I agree (see section 6, as well as Bishop and Trout [2005]). But this is not inconsistent with the *purely descriptive* claim I'm arguing for here. My ultimate view is that reasoning rationally about selective defection cases will require a *significant revision* in the way our reasoning competence works (see also Bishop 2000). To make this case, I will document people's resistance to using improper models and then argue that the best explanation for this resistance is that the FMP is inconsistent with the principles that make up our reasoning competence.

In selective defection studies, subjects are given a successful SPR and allowed to override its judgment. The consistent result of these studies is that subjects are *still* outperformed by the SPR, even when they're told that it has been shown to be more reliable than experts (Sawyer 1966, Leli and Filskov 1984). The Goldberg Rule is a successful unit weight model that predicts whether a psychiatric patient is neurotic or psychotic on the basis of a MMPI (Minnesota Multiphasic Personality Inventory) profile. It diagnoses a patient as neurotic if (L + Pa + Sc − Hy − Pt) < 45 (where L is a validity scale and the rest are personality scales of the MMPI), otherwise it diagnoses the patient as psychotic. In a particularly interesting selective defection study, one group of subjects was given the Goldberg Rule and a second group of subjects was given only "the numerical value of the formula for each profile and the optimum cutting score [45]." Both groups were told that "this formula would achieve approximately 70% accuracy and that it would be more accurate for extreme values than for values close to the cutting score" (Goldberg 1968, 493). The results are represented in figure 2.

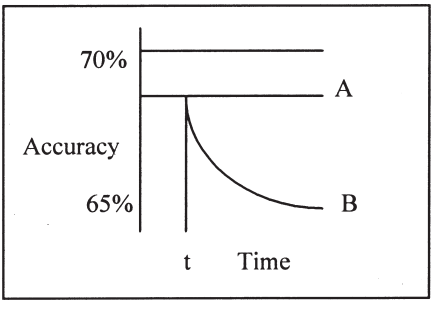


Figure 2: The Goldberg (1968) selective defection result

Those who were not given the formula but only its results (A) "increased their accuracy to a bit below 70% correct" – still not as accurate as the formula all by itself. For those who were given the formula but had to solve it themselves (B), there was a short-term increase in accuracy. But this improvement "gradually wore away over time" until it disappeared altogether (Goldberg 1968, 493). Group B ended up reasoning about as accurately as if they had never been given the SPR. Let's distinguish two aspects of this result.

1. Resistance: We defect from SPRs even when we know they are more reliable than we are in the long run.
2. Suboptimality: Experts with SPRs are less reliable in the long run than SPRs.

It is useful to keep these phenomena distinct. While some resistance is obviously necessary for suboptimality, it is not sufficient. We might resist and improve upon the SPR's reliability, which is what experts certainly thought they were doing. An obvious explanation for these results would appeal to our ignorance and poor reasoning. In particular, subjects are ignorant of the FMP, the SPR results, and they perhaps don't fully believe that the SPR is more reliable than they are; and when it comes to complex issues, we don't reason as reliably as we think we do. Although J.D. Trout and I have defended an explanation along these lines (Bishop and Trout 2005, 37-53), there are two prima facie problems with it:

a. It doesn't account for the difference between groups A and B in resistance and suboptimality. Are we supposed to believe that the members of group A know more about the FMP, SPRs or the Goldberg Rule than the members of group B, or that group A consists of better reasoners than group B? Surely not.
b. It doesn't account for the delay in group B's resistance and suboptimal reasoning.

While ignorance and poor reasoning are part of

 https://doi.org/10.3366/epsy.2005.2.1.65 Published online by Cambridge University Press

the explanation for the selective defection results, they can't be the full story.

A more complete explanation begins with the assumption that people are quite good, but not perfect, at tracking epistemic support (i.e., when the evidence supports one belief better than another). In figure 3, Expert has only the IM, not the PM. Since the PM is constructed so as to best fit our available evidence, we'll suppose that it always delivers the belief that is best supported by the evidence. (This assumption is usually, but not always, true; see section 4.)

	1	2	3	4	
	55%	*15%*	*15%*	*15%*	% Correct
Proper model (PM)	T	F	T	F	*70%*
Improper model (IM)	T	F	F	T	*70%*
Expert: Defection Rate from IM	*5%*	*5%*	*95%*	*95%*	68%

Figure 3: Expert tracks epistemic support 95% of the time

Let's assume that Expert tracks epistemic support 95% of the time.[4] As a result, she defects from IM *5%* of the time when it agrees with PM (columns 1 and 2) and she defects 95% of the time when it disagrees with PM (columns 3 and 4). With this pattern of defection, Expert loses 2% accuracy (*2.75%* loss in column 1, a *.75%* gain column 2, and no net change in columns 3 and 4 combined). But Expert comes to the belief that is best supported by the evidence (i.e., agrees with the PM) 95% of the time (as opposed to only *70%* for the IM). So Expert reasons less reliably than the IM even though she is *much* more likely to adopt the belief that is best supported by the evidence.[5]

An essential part of the correct explanation of the selective defection results must appeal to the fact that we track evidential support better than improper models. We typically defect from a SPR on the basis of a second-order belief to the effect that the SPR recommends a belief that is not best supported by the evidence. If figure 3 represents the reasoner's situation in a selective defection study, note two points about these second-order beliefs:

1. They are usually true (e.g., in figure 3, they're true 95% of the time).
2. They are typically supported by undefeated (and often valid) arguments.

Given (1) and (2) and almost any reasonable theory of justification, *when we defect from a SPR, we are justified in believing that the SPR recommends a belief that is not best supported by the evidence*. When this second-order belief is true (in columns 3 and 4) we can draw an even stronger conclusion: *We know, sometimes on the basis of a deductively sound argument, that the IM-recommended belief is not best supported by the evidence*. No wonder we defect.

But why do we reason suboptimally? The answer is that high sensitivity brings false alarms. We are like the overly sensitive fire alarm that reliably detects fires but also smoky meats. By being so sensitive to the evidence, we usually defect when the evidence doesn't support the belief recommended by the SPR (columns 3 and 4), but we also occasionally defect when the evidence does support the belief recommended by the SPR (columns 1 and 2). It is reasonable to suppose that ignorance and poor reasoning play a significant role in cases of inappropriate defection. The problem is that defecting when we "shouldn't" (columns 1 and 2) decreases our reliability whereas defecting when we "should" (columns 3 and 4) does not change our reliability.

This explanation has the resources to account for the pattern of defection in figure 2. Group B defects more often than group A simply because members of B more often come to the relevant second-order belief (that the evidence does not support the SPR recommended belief). And this is because group B's use of the SPR puts them in a better position to know when the SPR is ignoring or otherwise misweighing evidence. Consider a schematic example. Group B members (unlike group A members) know which MMPI scores the SPR ignores (or misweighs); so they are in a better position to detect cases in which those ignored (or misweighed) scores indicate a diagnosis that contradicts the SPR's diagnosis. When this happens, members of group B defect because they recognize that abiding by the SPR requires them to adopt a belief that is not supported by the evidence.[6]

https://doi.org/10.3366/epi.2005.2.1.65 Published online by Cambridge University Press

4. Everybody's Objection

In my experience, (almost) everybody responds to these studies as follows: "I'll accept the selective defection results and your explanation for them. But nothing here indicts our reasoning competence. Subjects reason suboptimally in the selective defection studies because they fall victim to a performance error. They fail to give sufficient weight to an important piece of evidence, namely, that the SPR is in the long run more reliable than they are. If they were to weigh this evidence properly, they wouldn't reason less reliably than the SPR." In thinking clearly about Everybody's Objection, it's important to keep in mind that the issue is not whether subjects in the selective defection studies are somehow irrational. (I am on record arguing that they are.) The issue is whether their reasoning in the selective defection studies is consistent with the principles that constitute our reasoning competence or whether it is the result of a performance error (ignorance or a false belief or a failure of an ancillary cognitive system).

It is easy to boldly assert Everybody's Objection, confident that your reasoning competence would direct you to reason as reliably as the SPR. But in constructing her epistemological theories, the individualist does not rely primarily on these sorts of abstract judgments but on our epistemic intuitions about particular, well-defined cases (like the Gettier examples). Everybody's Objection needs to be tested against some concrete, albeit hypothetical, cases – preferably cases involving subjects about which you have some expertise. If you are a philosopher with some experience on hiring committees, you probably consider yourself a reasonably good judge of philosophical promise. So let's suppose there are two SPRs for hiring philosophers that have been shown to be more reliable in the long run than experts (yes, you).[7] One is a proper model (PM) that considers all the evidence you normally consider in making hiring decisions. For the sake of convenience, let's suppose this involves four lines of evidence: letters of recommendation, Ph.D. program, writing sample, and interview. The second SPR is an improper model (IM) that weighs three lines of evidence equally (so it is a unit model) and ignores the least predictive line of evidence – let's say it ignores the interview evidence.

We'll consider four hypothetical selective defection cases to test your epistemic intuitions. In every case, we'll assume that you only have the IM available, and that you and the IM agree about who the two finalists are – Theresa (T) and Fabiola (F). Since you know the FMP and the SPR results, we'll suppose you have well-grounded confidence that figure 4 represents your situation reasonably accurately.

	1	2	3	4	
	55%	*15%*	*15%*	*15%*	% Correct
Proper model (PM)	T	F	T	F	*70%*
Improper model (IM)	T	F	F	T	*70%*
Your prospective defection rate	*5%*	*5%*	95%	95%	68%

Figure 4: The hiring cases

So in these four cases, you will have two epistemological considerations pulling against each other. On the one hand, considering only the evidence you have about the candidates, you will have a justified belief that the evidence supports defecting from the SPR. On the other hand, the SPR recommends a higher frequency of true beliefs in the long run than experts like you. How strong is the frequency consideration? In figure 4, you are inclined to defect 32% of the time (by column: 2.75% + .75% + 14.25% + 14.25%); and you will defect to the true belief 15% of the time and to the false belief 17% of the time. Not a huge difference. *If you defect, you can expect to defect to the true belief almost 47% of the time.*[8, 9] Keep in mind, this statistic is a long run frequency; it says nothing about whether you're right to defect in any particular case.

> Case #1: You have applied the IM to the dossiers of T and F. This has involved grading the applicants' letters, writing samples, and Ph.D. programs, normalizing those grades (according to simple functions), and then adding those transformed grades together. F has a slightly higher total grade than T, and so the IM recommends hiring F. You are fairly sure, however, that the PM would recommend hiring T, and that the entirety of the evidence supports hiring T. The reason is that the IM

70 https://doi.org/10.3366/epd.2005.2.1.65 Published online by Cambridge University Press

ignores interview information. T gave a terrific interview. However, just after F's interview began, F verbally abused and then physically assaulted the interviewers.

If you defended Everybody's Objection because you were confident that you would *never* defect from a successful SPR, you were kidding yourself. The fact that one line of evidence (e.g., interviews) is not strongly correlated with job performance is a fact about long run frequencies. On any particular occasion, that evidence might be highly relevant. But this will be infrequent enough that it won't make a significant difference to the SPR's long term reliability.

Case #2: You have applied the IM to the dossiers of T and F. F has a higher total grade than T, and so the IM recommends hiring F. You are fairly sure that the PM would also recommend hiring F. However, you have extra evidence about F that is not usually considered in hiring. At the national conference, you observed F (but no other candidates) after hours. F was engaged in shocking and illegal behavior – walking down the street with a weapon and a bottle of whiskey, yelling, "I'm going to bag me some liberals tonight!"

The first two cases show that *SPRs, including proper models, ignore relevant evidence*. The ignored evidence is relevant in the sense that one could in principle improve on the SPR's long term reliability by taking it into account. Of course, cases #1 and #2 are extreme and unrealistic, but then so are many important examples that inform contemporary epistemology. More importantly, the staunchest defenders of SPRs recognize that SPRs ignore relevant evidence and that defection is sometimes justified.[10] A psychologist who developed a SPR for predicting recidivism told me of a striking example of legitimate defection: one should always correct his model if it doesn't predict a zero chance of recidivism for dead people.[11]

At this point, Everybody's Objection becomes: "Ok, my reasoning competence would direct me to sometimes defect from a successful SPR. But I wouldn't resist as much as subjects do in selective defection tasks, because I would properly weigh the fact that the SPR is more reliable in the long run than I am." The problem, of course, is that instances of legitimate defection don't announce themselves with fanfare and confetti.

Case #3: You have applied the IM to the dossiers of T and F. Comparing the dossiers, T and F went to the same university (so the program quality is the same), T's writing sample is inferior to F's, but T's letters are superior to F's. But the difference in the quality of the writing samples is appreciably greater than the difference in the quality of the letters. Since the IM weighs the three pieces of evidence equally, it ranks F higher than T. But given your experience, you justifiably believe that letters are stronger evidence of future job performance than the writing sample; if one had to make a decision based on only one of these, one would choose the letters. Had the model given somewhat greater weight to the letters than to the writing sample, as any proper model would, you are convinced that it would have recommended hiring T.

You have two powerful arguments for defection:

(a) The quality of the letters is more strongly correlated with job performance than the quality of the writing sample; and so F's superior writing sample is less diagnostic of job performance than T's superior letters. The evidence best supports the belief that T has the stronger credentials.
(b) The improper model comes to the opposite conclusion because it does not weigh these lines of evidence appropriately. If it did, it would have come to the conclusion that T has stronger credentials than F.

Of course, you know that when you defect, your long term accuracy is slightly worse than the SPR's (about 47% to 53%). But you also know there are legitimate cases for defection. When you defect, you *know* that T has stronger credentials than F 89% of the time.[12, 13] So on just about anybody's view of justification, it would seem that you are justified in believing that T has stronger credentials than F. (And if T really does

https://doi.org/10.3366/epi.2005.2.1.65 Published online by Cambridge University Press

have stronger credentials, you *know* this.) In this epistemic situation, what do your epistemic intuitions – your reflective, non-discursive judgments – tell you about what you ought to believe?[14]

Let's consider one final case.

> Case #4: You have applied the IM to the dossiers of T and F. The IM ranks F somewhat higher than T. But IM ignores interview evidence. As it happens, T had a *much* better interview than F. In fact, T sparkled, while F appeared almost incompetent, sometimes unable to answer questions coherently. You have no objection to the IM recommendation given only the evidence the IM considered. But you are certain that if the IM were amended so as to give even a slight degree of weight to the interview evidence, the large difference in the quality of their interviews would lead the model to recommend hiring T rather than F.

You have two powerful arguments for defecting from the IM:

(a) While interviews are only weakly correlated with job performance, in this case, T and F are sufficiently closely matched on other measures that the difference in their interviews tips the scale in T's favor. The evidence best supports the belief that T has the stronger credentials.
(b) The improper model comes to the opposite conclusion because it completely ignores interview evidence. If it weighed interview evidence appropriately, it would have come to the conclusion that T has stronger credentials than F.

Once again, you know that when you defect, your long term accuracy is slightly worse than the SPR's. But you also know that there are legitimate cases for defection. In this case, you are justified in believing that T has stronger credentials than F. What do your epistemic intuitions – your reflective, non-discursive judgments – tell you about what you ought to believe?

Those who are inclined to defect in cases #3 or #4 might object that I have rigged the game by describing cases in which one has strong evidence in favor of defecting. If you are inclined to raise this objection, I have to wonder: What were you expecting? That expert clinicians with years of experience would defect from SPRs arbitrarily or when they had clearly insufficient reasons? Such assumptions are uncharitable. Further, they betray a failure to appreciate the deep weirdness of the implications of the flat maximum principle. In selective defection tasks, the FMP directs us to ignore and misweigh relevant evidence when making highly significant judgments concerning matters about which we have considerable expertise. Finding that hard? Join the club.

What about those who insist that in cases #3 and #4, their epistemic intuitions are telling them not to defect? There is a way to test your confidence. Many studies have found that people's performance on short, unstructured interviews is positively correlated with job performance; so interview evidence is relevant to predictions of job performance. But we think this correlation is much stronger than it really is. Actual correlations are typically less than .10, whereas Kunda and Nisbett found that subjects' average estimate of the correlation was .59 (1986, 213). The "interview effect" is a result of this overconfidence: when we take interview performance into account, we tend to make worse gatekeeping judgments because lesser quality (interview) information tends to swamp higher quality (dossier) information (Hunter & Hunter 1984).[15] So the next time your department hires, ignore all the conference interview information.[16] At some point, you will find yourself in a situation analogous to case #4 – arguing for someone who bombed the interview and against someone who sparkled. In this case, would your intuitive judgment be that you should ignore the interview evidence? And assuming you clearly state the case for ignoring the relevant interview evidence, what would be your colleagues' intuitive judgments? My guess is that your confidence in the ability of our reasoning competence to ignore relevant evidence won't survive this empirical scrutiny.

Now suppose I'm wrong. You and your department pass this test with flying colors. You ignore the interview evidence with impunity. But after a time, suppose it becomes obvious that hiring the chap who bombed the interview was a

 https://doi.org/10.3366/epd.2005.2.1.65 Published online by Cambridge University Press

mistake. (Given the assumptions in figure 4, non-defectors can expect to be wrong almost half the time.) We have epithets for people who reason to a false belief about a significant matter because they intentionally ignore relevant evidence. But the final test isn't whether you or your colleagues would level accusations of irrationality (or worse), but whether over time, your determination to ignore relevant evidence would wane. After having been burned by ignoring evidence you *knew* was relevant, the next time your department hires would you be less inclined to ignore evidence you were justified in believing was relevant?[17] Consider again the resistance finding represented in figure 2. Group B doesn't resist much at first; it's only later, after they've had experience with the SPR that they really resist ignoring relevant evidence. So being able to ignore relevant evidence once, or even on a few occasions, isn't enough. You must remain determined to ignore evidence you know is relevant, even after you've been burned by doing so. And the longer you keep at it, the more you can expect to be burned by it. If you are inclined to lose patience with making errors as the result of ignoring evidence you knew was relevant, then you are inclined to exhibit a classic pattern of resistance.

5. Explaining Excessive Defection

It is time to confront the consideration that I think ultimately drives Everybody's Objection: "Surely, there is *some* mistake that we are making when we defect excessively from successful SPRs. After all, this defection leads us to reason less reliably than the SPR. Once we put our finger on this mistake, it will be obvious that it is a performance error and not a result of the proper operation of our reasoning competence." This consideration will continue to drive Everybody's Objection unless I can explain, even tentatively, how the selective defection results might be the result of the normal operation of our reasoning competence, rather than the result of ignorance or false beliefs or bad reasoning. The crux of my speculative explanation is that our reasoning competence takes non-defection errors to be epistemically worse than defection errors. Our reasoning competence operates in such a way that it is prepared to make more total errors in order to make fewer non-defection errors. It is important to keep in mind that the outputs of our reasoning competence are epistemic intuitions – reflective but non-discursive epistemic judgments. So our disposition to accept this trade-off (more errors for fewer non-defection errors) is no more the result of conscious deliberation than are our Gettier judgments.

Why might our reasoning competence take non-defection errors to be worse than defection errors? Here are two mutually supporting hypotheses. First, psychologists have shown that regret plays an important role in our judgment and decisionmaking. We know we will feel regret if we make an error. We take this anticipated regret into account when making decisions. So ceteris paribus, we avoid choices that we think will bring greater regret if they turn out badly (Loomes & Sugden 1982). Perhaps our epistemic intuitions in the selective defection studies are driven in part by the fact that we anticipate greater regret from non-defection errors than from defection errors. Consider the regret you would feel if you made a bad hire and your mistake was "We didn't pay attention to what we knew the evidence was saying" as opposed to "We defected from a formula that we knew was about 2% more reliable than we are in the long run." It is not implausible to suppose we anticipate considerably greater regret with the former mistake than the latter. If so, then perhaps our reasoning competence reflects this fact when guiding how we reason.

The second hypothesis is suggested by a parole board commissioner who explains why he refuses to use SPRs: "If you had the ability to look at someone and make a, you know, conscientious decision, but you chose to look at a piece of paper instead and made the wrong thing and maybe you could have allowed somebody to spend some more time with their children, you know, it's going to burden you forever" (Spiegel 2004). The commissioner apparently feels less burdened by the fact he's making a greater number of equally consequential errors as a result of ignoring the SPR. Perhaps we find non-defection errors epistemically worse than defection errors because we feel more epistemically responsible for non-defection errors than for defection errors. We feel greater responsibility when we adopt a belief we

https://doi.org/10.3366/epi.2005.2.1.65 Published online by Cambridge University Press

foresaw to be an error, an error we could have prevented but didn't, than when we adopt a belief about p that we were justified in thinking was best supported by the evidence we had about p. Many epistemologists have argued that our notion of epistemic justification is tied to the notion of epistemic responsibility (e.g., Kornblith 1983, Plantinga 1993). If so, then we would expect our epistemic intuitions, if they track justification reasonably well, to incline us away from non-defection errors.

My suggestion is that our suboptimal reasoning in the selective defection studies is the not result of ignorance or false beliefs or poor reasoning. It is the result of the proper operation of our reasoning competence, which directs us to accept more total errors in order to make fewer non-defection errors.

6. Objections and Implications

Social institutions have no trouble abiding by the FMP. A financial institution might make thousands of credit decisions a day. A hospital might make thousands of medical diagnoses a year. From the perspective of these institutions, error is inevitable. What matters in making these thousands of judgments is minimizing significant error. So if institutional decision-makers are faced with using either a SPR that minimizes significant error or human experts who make considerably more significant errors, the choice is a no-brainer. The fact that the experts would make fewer non-defection errors cuts no ice.

I believe that our reasoning competence can be reformed to better handle selective defection cases. In this spirit, the individualist can (and, I believe, should) co-opt the FMP as a principle of individual rationality. This will be particularly easy if she accepts the method of wide reflective equilibrium (Goodman 1965, Daniels 1979). But however the individualist goes about justifying the revision of our notion of individual rationality in the light of the FMP, isn't there something disingenuous about the individualist absorbing the FMP while denying the autonomy of social epistemology? After all, the individualist has to revise our tacitly held principles of individual rationality because they are *mistaken*: they are practically inconsistent with a correct principle of rationality, the FMP. So when the principles of our reasoning competence are right, they're right; and when they're wrong, the individualist revises them without guilt or penalty. Nice racket.

An individualistic theory of rationality that simply articulates and clarifies the principles of our reasoning competence will end up being incorrect. Such a theory must be informed and improved by principles of social rationality – principles governing epistemically well-ordered social systems. It's perfectly reasonable for the individualist to turn around and absorb elements of social epistemology. But it would be delusional for him to then insist that social epistemology does not have its own, distinctive theory – that it simply borrows the principles of individual epistemology and applies them to cases in which social factors are crucial.

While reliabilists are likely to be sympathetic to the argument presented here, they might not take it as a case for autonomy. They might argue that the autonomy thesis is false because social epistemology does not have a unique set of basic normative principles; one normative theory, reliabilism, is all we need for both individual and social epistemology. But this objection fails to appreciate the logic of the case for autonomy. The case depends essentially on the fact that reliabilism is inconsistent with the principles that constitute our reasoning competence. So reliabilism cannot be a particularly good theory of individual rationality (where the goal, recall, is to provide a theory that captures our epistemic intuitions). Of course, the reliabilist is free join the other individualists in insisting that we should *revise* our reasoning competence and our epistemic intuitions in light of what we have learned about the principles that govern epistemically well-ordered social systems. But then she is in the same predicament as the individualist in the last two paragraphs: correcting our principles of individual rationality with the principles of social rationality. If we take reliabilism to be a promising theory of social epistemology rather than a doomed theory of individual rationality, then reliabilists have always embraced the autonomy thesis, whether they've known it or not.

Some might be disappointed that the conception of social epistemology I have

 https://doi.org/10.3366/epd.2005.2.1.65 Published online by Cambridge University Press

employed in this paper is a rather conservative, reliabilist one. For the disappointed, let me emphasize two points. First, my goal has been to argue that even on very conventional, individualism-friendly conceptions of rationality, social epistemology is autonomous. Nothing in this paper commits anyone to these assumptions. And second, the autonomy thesis I have defended isn't especially modest. Let me briefly (and without much argument) mention three ways in which the autonomy thesis is bolder than it might first appear.

1. The autonomy of social epistemology does not mean that it has no connections to other disciplines. It is likely to depend crucially on findings in the social sciences – psychology, economics, history and sociology. But armed with pragmatic and semantic notions (e.g., useful truth), social epistemology is unlikely to need any complicated results from traditional, individual epistemology. Good social epistemologists can (and often do) ignore traditional, individual epistemology as it has been practiced since Plato.
2. The case for autonomy raises worries about the prospects of methodological individualism, which is practiced by most contemporary epistemologists. If our reasoning competence can conflict with correct epistemological principles, then it seems perverse to insist upon epistemological theories that capture the judgments of our reasoning competence. This worry is magnified by evidence showing that people's reasoning competence seems to vary by culture and socioeconomic class (Weinberg, Nichols & Stich 2001).
3. Social epistemology is autonomous because there is some inferential knowledge that derives uniquely from the principles of social rationality. These principles can be co-opted by individualists, but their home is in social epistemology. But then, why not suppose that *all* inferential knowledge derives from the principles of social rationality? In order to find a legitimate, non-redundant role for individual rationality, the individualist must undertake the treacherous task of drawing a principled distinction between social (inferential) knowledge and individual (inferential) knowledge. The individualist must carve out a privileged domain of inferential knowledge where only individual rationality applies. If (as I suspect) this is not possible, then individual rationality is redundant.

Although we began with individualist-friendly conceptions of rationality, the autonomy thesis suggests that the commonsense, individualist view has it exactly backwards. The legitimacy and autonomy of social epistemology cannot be reasonably questioned. But there is reason to believe that individual rationality is otiose. In that case, social epistemology is autonomous because social rationality *is* rationality.

Turning epistemology on its head in this way helps make sense of the pronounced tendency among contemporary epistemologists to offer theories that individuals have no hope of implementing. An individual untrained in philosophy, working all by herself, could probably not understand, much less effectively employ, most contemporary epistemological theories to make significant, systematic improvements in how she reasons. But social institutions can – and do. By turning epistemology on its head, we no longer need to take these epistemological theories to be unrealistic or outrageously idealized theories of individual rationality. They are perfectly implementable theories of social epistemology that are being applied (perhaps badly applied) to the individual. Epistemology has always been social.

References

Ashenfelter, O., D. Ashmore, & R. Lalonde (1995). Bordeaux Wine Vintage Quality and the Weather. *Chance* 8: 7-14.

Bealer, G. (1987). The Philosophical Limits of Scientific Essentialism. *Philosophical Perspectives* 1: 289-365.

Bishop, M. (2000). In Praise of Epistemic Irresponsibility: How Lazy and Ignorant Can You Be? *Synthese* 122: 179-208.

https://doi.org/10.3366/epi.2005.2.1.65 Published online by Cambridge University Press

Bishop, M. & J.D. Trout (2005). *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press.
BonJour, L. (1998). *In Defense of Pure Reason: A Rationalist Account of A Priori Justification*. London: Cambridge University Press.
Carroll, J., R. Winer, D. Coates, J. Galegher, & J. Alibrio (1988). Evaluation, Diagnosis, and Prediction in Parole Decision-Making. *Law and Society Review* 17: 199-228.
Cohen, L.J. (1981). Can Human Irrationality be Experimentally Demonstrated? *Behavioral and Brain Sciences* 4: 317-331.
Daniels, N. (1979). Wide Reflective Equilibrium and Theory Acceptance in Ethics. *Journal of Philosophy* 76: 256-82.
Dawes, R. (1982). The Robust Beauty of Improper Linear Models in Decision- Making In *Judgment under Uncertainty: Heuristics and Biases,* eds. D. Kahneman, P. Slovic & A. Tversky, 391-407. Cambridge: Cambridge University Press. First appeared in (1979) *American Psychologist* 34: 571-582.
Dawes, R. (1988). *Rational Choice in an Uncertain World*. New York: Harcourt.
Dawes, R. & B. Corrigan (1974) Linear Models in Decision Making. *Psychological Bulletin* 81: 95-106.
Dawes, R., D. Faust, & P. Meehl (1989). Clinical Versus Actuarial Judgment. *Science* 243: 1668-1674.
Einhorn, H. J. & R.M. Hogarth (1975). Unit Weighting Schemas for Decision Making. *Organizational Behavior and Human Performance* 13: 172-192.
Faust, D. & J. Ziskin (1988). The Expert Witness in Psychology and Psychiatry. *Science* 241: 1143-1144.
Goldberg, L.R. (1968). Simple Models of Simple Processes? Some Research on Clinical Judgments. *American Psychologist* 23: 483-496.
Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: MIT Press.
Goldman, A. (1999). *Knowledge in a Social World*. Oxford: Oxford University Press.
Goldman, A. & J. Pust (1994). Philosophical Theory and Intuitional Evidence In Schmitt, F.F. (ed) *Socializing Epistemology: The Social Dimensions of Knowledge*. Lanham, MD: Rowman & Littlefield Publishers, Inc.
Goodman, N. (1965). *Fact, Fiction, and Forecast*. Indianapolis, IN: The Bobbs-Merrill Co., Inc.
Hunter, J.E. & R.F. Hunter (1984). Validity and utility of alternative predictors of job performance *Psychological Bulletin*, 96: 72-98.
Kitcher, P. (1990). The Division of Cognitive Labor. *Journal of Philosophy* 87: 5-21.
Kornblith, H. (1983). Justified Belief and Epistemically Responsible Action. *The Philosophical Review* 92: 33-48.
Kunda, Z. & R.E. Nisbett (1986). The Psychometrics of Everyday Life. *Cognitive Psychology*, 18: 195-224.
Leli, D.A. & S.B. Filskov (1984). Clinical Detection of Intellectual Deterioration Associated With Brain Damage. *Journal of Clinical Psychology* 40: 1435-1441.
Loomes, G. & R. Sugden (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal, 92*, 805-824.
Meehl, P. (1986). Causes and Effects of My Disturbing Little Book. *Journal of Personality Assessment* 50: 370-375.
Passell, P. (1990). Wine Equation Puts Some Noses Out of Joint. *The New York Times,* March 4: 1.
Plantinga, A. (1993). *Warrant: The Current Debate*. New York: Oxford University Press.
Pust, J. (2000). *Intuitions as Evidence*. New York: Garland Publishing.
Sawyer, J. (1966). Measurement and Prediction, Clinical and Statistical. *Psychological Bulletin* 66: 178-200.
Schmitt, F. (1994). The Justification of Group Beliefs In Schmitt, F.F. (ed) *Socializing Epistemology:*

https://doi.org/10.3366/epi.2005.2.1.65 Published online by Cambridge University Press

The Social Dimensions of Knowledge. Lanham, MD: Rowman & Littlefield Publishers, Inc.
Solomon, M. (2001). *Social Empiricism*. Cambridge, MA: MIT Press.
Spiegel, Alix (2004). Profile: Comparison of human intuition to statistical analysis for predicting future human behavior. *All Things Considered*. National Public Radio, 5 January 2004. Broadcast.
Stillwell, W., F. Barron, & W. Edwards (1983). Evaluating Credit Applications: A Validation of Multiattribute Utility Weight Elicitation Techniques. *Organizational Behavior and Human Performance* 32: 87-108.
Wedding, D. (1983). Clinical and Statistical Prediction in Neuropsychology. *Clinical Neuropsychology* 5: 49-55.
Weinberg, J., S. Nichols, & S. Stich (2001) Normativity and Epistemic Intuitions. *Philosophical Topics*, 29 (1 and 2): 429-460.

Notes

[1] For helpful comments on earlier versions of this paper, I want to thank Jennifer Lackey, Baron Reed, J.D. Trout and the audience at the Second Annual EPISTEME Conference, particularly Dennis Whitcomb, David Christensen, Don Fallis, Alvin Goldman, Gareth Matthews, Gurpreet Rattan, Harvey Siegel, and Jonathan Vogel. This research was supported by a National Science Foundation grant (SES#0354536).

[2] Performance errors include errors that arise from the failure of ancillary cognitive systems (which is the standard notion of performance error taken from linguistics) and errors that are the result of ignorance or of false background beliefs.

[3] The qualification 'systematic' recognizes that most methodological individualists will permit some conservative revision of the principles that constitute our reasoning competence in the service of coherence and clarification.

[4] I will assume that we are very good trackers of epistemic support. While I worry that this is too optimistic, it is crucial for the methodological individualist. Why would she build epistemological theories that capture our epistemic intuitions if she thought we're lousy at tracking important epistemic properties?

[5] I chose 70% reliability scores for the IM and PM because that is the reliability score of the Goldberg Rule. But the distribution of agreement between the IM and PM (55/15/15/15) is purely hypothetical. The claim made here is insensitive to this distribution. As long as the IM and PM have 70% reliability scores, a 95% tracker of PM's judgments will lose 2% reliability.

[6] The delay occurs because members of group B have to undergo a process of coming to recognize that the SPR is improper and that it sometimes recommends beliefs that are not supported by the evidence. And then they have to be motivated to defect (more often than group A) in the face of the knowledge that the SPR is more reliable than they are. This process takes time.

[7] Perhaps you are skeptical that such a model could exist. This is to be expected. Recall the wine-tasters' reaction to the successful wine prediction SPR – "somewhere between violent and hysterical."

[8] If you track epistemic support 80% of the time, you'll defect to the true belief a bit less than 40% of the time. If you track epistemic support much worse than this, one wonders: Why does the individualist seek epistemological theories that capture our epistemic intuitions if we track epistemic support so poorly?

[9] The pattern of agreement between IM and PM makes a difference here. If we suppose they agree 90% of the time, so that the pattern of agreement is (65/25/5/5) rather than (55/15/15/15), then you defect 14% of the time and you defect to the true belief 6% of the time. So if you defect, you can expect to defect to the true belief almost 43% of the time. (Even if they always agree,

you'll defect to the true belief 30% of the time.)

[10] Here is Dawes, Faust and Meehl (1989, 1670): "Clinicians might be able to gain an advantage by recognizing rare events that are not included in the [SPR] (due to their infrequency) and that countervail the [SPR's] conclusion… The clinician may beat the [SPR] if able to detect the rare fact and decide accordingly. In theory, [SPRs] can accommodate rare occurrences, but the practical obstacles are daunting… [T]he possible range of intervening events is infinite."

[11] I will ignore the fact that proper models don't always deliver the belief that is best supported by the evidence, even though it strengthens my argument. A pillar of my argument is that we reason less reliably than IMs even though we track evidential support better than IMs. In assessing how well we track evidential support, I assume throughout this paper that the PM-recommended belief is best supported by the evidence. Since this will not always be true, I am *consistently underestimating* how often we adopt the belief that is best supported by the evidence.

[12] You defect 32% of the time. You defect to the PM-recommended belief in columns 3 and 4 (28.5%) and away from the PM-recommended belief in columns 1 and 2 (3.5%). So when you defect on the basis of a second-order belief that the IM-recommended belief is not best supported by the evidence, that belief is true 89% of the time.

[13] The distribution of agreement between IM and PM (55/15/15/15) makes a fairly significant difference to this calculation. Suppose IM and PM agree almost always, say, 90% of the time (65/25/5/5). As a 95% tracker, you'll defect to the PM-recommended belief 9.5% of the time, and away from it 4.5% of the time. So when you defect, you'll know that T has stronger credentials than F almost 68% of the time.

[14] No fair whipping out Bayes theorem. Whatever judgment you end up with as a result of applying Bayes' theorem, it won't be a non-discursive one. Methodological individualists build theories on the basis of our epistemic intuitions – our non-discursive judgments – about cases like these.

[15] Many of you are denying that you fall victim to the interview effect. Standard considerations include: (a) interviews have saved us from poor hires; (b) we have made very good hires using interviews; (c) interviews give us good evidence about a candidate's teaching ability; (d) interviews give us good evidence about a person's character (and whether she will "fit" in with the department). Ignoring the fact that there is direct evidence against some of these considerations, none of these are good arguments unless you have an appropriate control: In the long run, would your hires have been better or worse without the interviews? I have found that when academics – including scientists – are confronted with the fact that their argument is poor because their experience lacks the necessary controls (and in most cases a respectable sample size), a not uncommon reaction involves granting the point but remaining confident in one's ability to properly sift interview evidence.

[16] If you doubt whether this finding applies to hiring philosophers, consider the anecdotal evidence that is beginning to pile up: there are philosophy departments – from top Ivy League graduate programs to fine undergraduate programs at state universities – that have abjured convention interviews for years now and are very satisfied with the results.

[17] Before you argue that it is perfectly rational to modify a hiring practice when it fails, ask yourself: Would you support (or have you ever supported) radical changes to your department's hiring practices after a single poor hire?

Michael A Bishop is Associate Professor of Philosophy at Northern Illinois University. He has authored or coauthored over 20 articles on a wide range of issues in philosophy of science and epistemology. Issues tackled in these articles include: the nature and role of thought experiments in science, the virtues of epistemic irresponsibility, naturalism in epistemology, theory-ladenness of perception, semantic incommensurability, and the scientific realism-antirealism debate. Bishop is co-author, with J.D. Trout, of *Epistemology and the Psychology of Human Judgment* (2005).

78 https://doi.org/10.3366/epsy.2005.2.1.65 Published online by Cambridge University Press