Smart People's Rational Mistakes

Nathan Berg

University of Otago (Nathan.Berg@otago.ac.nz)

> [E]very intelligent system makes good errors; otherwise it would not be intelligent. The reason is that the outside world is uncertain, and the system has to make intelligent inferences based on assumed ecological structures. Going beyond the information given by making inferences will produce systematic errors. Not making these errors would destroy intelligence.
>
> – Gerd Gigerenzer (2005, p. 199)

I will describe theoretical and empirical examples of errors—both in games against nature and in strategic settings—that confer individual-level and, in some cases, Pareto-improving benefits to an entire economy or social system. My goal is to demonstrate the wide range of mechanisms by which we individually and collectively benefit from behaviors that behavioral economists have been too quick to label as mistakes, simply because those behaviors do not conform to the orthodox rational choice standard of rationality based on internal logical consistency. I want to invite you to reconsider the interpretations that can and should be attached to behavioural patterns that are commonly described by behavioral economists as decision-making errors.

I will argue that mistakes are vital for strengthening and maintaining valuable relationships and enabling perceptual, inferential, social and financial success. The implication is that smart people *must* (i.e., descriptively, as a logical consequence of the requirements of success) and *should* (i.e., prescriptively) make mistakes.

How could success require mistakes? Are such notions of `mistakes' merely a semantic parlour trick that disappears once proper definitions of success are introduced? Does the claim that smart people must make mistakes sound like a bad joke? In fact, the mistake of telling a `bad joke' (i.e., an ill-chosen attempt at humor that unintendedly winds up annoying or offending someone you care about and want to feel good) illustrates the point precisely that smart people must err (c.f., the examples and arguments in Gigerenzer, 2005).

Bad Joke

Consider what happens if I tell my girlfriend a story that I heard from an entertaining and rough-talking (read "severely politically *in*correct") friend of mine. It turns out that my girlfriend does not like the joke and finds it deeply offensive. Normally, she would update her beliefs about any person (e.g., me) who has spoken those specific words aloud.

My girlfriend might, in fact, be interpreted here as a Bayesian updater whose subjective belief that I am a high-quality, worthy person (after conditioning on the historical sequence of speech acts by me that she has observed). Normally, her conditional assessment of my value would decline sharply conditional on my telling of the bad joke. The joke was so bad that, conditional on observing me tell it, her updating function abruptly downgrades the level of subjective esteem associated with the speaker (given the *finite* number of virtuous acts she has previously observed in our shared history).

I am assuming that she also uses a threshold condition to accept men worthy of dating (i.e., satisficing on mate choice while updating beliefs according to Bayes Law). Her updated level of esteem for me conditional on the bad joke now falls strictly below her minimum threshold required for mate acceptance. She would normally reject me out of hand as a partner based on her usual belief-updating system. The bad joke I told would therefore normally exclude me from her consideration set and lead to a break-up of the relationship already in progress. What should I expect comes next?

Instead of breaking up, she decides to forgive me. She says, "I didn't like the words I heard you say, but I forgive you. Please don't say that again."

And just like that, something outside the normative performance metrics introduced in the model so far newly enters the analysis. Like bones that heal stronger than in their previously unbroken state, or an immune system that heals stronger, apparently relationships, too, can grow deeper, richer, more valuable and stronger—in love, business, science, and in friendships of many kinds—thanks to the event of a mistake followed by forgiveness (or other means of relationship repair). I consider several modeling strategies with the goal of representing the mechanism of relationships whose depth or robustness benefits from mistakes and shared adversity in their shared history.

Among my reasons for raising this example are the subtleties it raises regarding the methodological conventions of constrained optimization and game-theoretic reasoning that behavioral economists typically use as the benchmark of perfect rationality in relation to which deviations are thought to measure irrationality, *a*-rationality, or various normative gradations of irrationalities (Berg, 2003, 2014a; Berg and Gigerenzer, 2006, 2010). Next I will present contrasting representations of this interaction corresponding to different views of other players' action sets, whether those action sets include the possibility of intentionally-told versus randomly-told bad jokes (as a result of "nature's move"); and whether the continuation value of the relationship itself is included explicitly, possibly strengthened as a result of withstanding a threatening event and then recovering thanks to forgiveness and repair.

Small world with no intentionally bad jokes and no control over bad-joke probability p

Figure 1 shows a simple, small world with no possibility of intentionally telling a bad joke. Bad jokes are modelled in Figure 1 as a random event detached from any other variable under the joke teller's control such as effort. Note, too, that there is no explicit model of risk preferences, cautionary motives, pro-social affections or anti-social motives such as spite. The only decision variable that the joke teller, referred to throughout as Agent 1, has to make in Figure 1 is whether to tell a joke or not.

The second player whose payoffs are represented in Figure 1 is Agent 2, the receiver of Agent 1's joke (e.g., my girlfriend in the discussion above). Without loss of generality, both players' payoffs are normalized to zero at the left-most no-joke node, so that the payoffs associated with the other two terminal nodes, corresponding to bad- and good-joke outcomes, respectively, represent changes in payoffs relative to the (0, 0) no-joke outcome.

I assume that both players prefer the good-joke outcome, which implies that $b_1<g_1$ and $b_2<g_2$. Note that if $b_1>0$, then there is no downside risk in telling the bad joke (relative to telling no joke) because the joke teller's bad-joke payoff is strictly greater than zero. To make the risk of landing on the bad-joke outcome interesting, the bad-joke payoffs should be negative numbers.

The probability that nature draws a bad-joke outcome once Agent 1 has decided to tell a joke is p, $0<p<1$. I further assume that the joke teller, Agent 1, is an expected payoff maximizer, so that his decision to tell a joke is, ex ante, rationalized by the following (obvious) condition describing precisely when Agent 1's expected payoff from joke telling is positive, $0 < pb_1+(1-p)g_1$, or, equivalently:

$$p < g_1/(g_1-b_1) \qquad \text{[condition rationalizing telling a joke in Figure 1].}$$

The condition above requires that the marginal gain of a good-joke relative to a no-joke outcome (g_1) as a fraction of the marginal gain of a good-joke outcome relative to a bad-joke outcome (g_1-b_1) is greater than the probability of failure (p) in order to rationalize telling the joke. Notice that this condition becomes non-binding if the ratio on the right-hand-side is strictly greater than 1, which occurs whenever $b_1>0$ as discussed above (because there is no loss in bad-joke relative to no-joke outcomes).

Assuming from now on that $b_1<0$ and $b_2<0$, should we then agree with conventional wisdom in behavioral economics that the bad-joke outcome is always best avoided if possible? In Figure 1, there is nothing in either player's choice set that enables him or her to control the probability of the bad-joke outcome. But if there were, would rationality then trivially require (i.e., by definition) that agents avoid making the mistake of telling a bad joke? The next model gives Agent 1 clairvoyance to focus on the question of whether he would ever rationally choose a bad joke he has perfect control to avoid.

Agent 1 has clairvoyance and therefore perfect ability to avoid telling the bad joke

In Figure 2, Agent 1 is clairvoyant or, equivalently, nature moves first and in a manner that is visible to both players determines the quality of the joke before Agent 1 decides whether to tell it. Therefore, Agent 1 knows in advance if the joke will land in Agent 2's ears as a bad one or good one.

An own-payoff-maximizing Agent 1 will never choose to tell a bad joke in the model depicted in Figure 2. If a bad joke occurs, then, because the joke teller is clairvoyant, Agent 2 knows that Agent 1 actually intended harm or offence. The bad-joke outcome in Figure 2 can never be accidental. Therefore, the possibility of spite or malevolence is now an unavoidable consideration for Agent 2 upon observing the bad joke. The unnatural abstraction of the payoffs from the context of the Agents' relationship shows up starkly and reflects a razor-edge view of what can be rational in Figures 1 and 2. The missing context of relationship remains in the next representation, which returns to the setup in Figure 1 but endows Agent 1 (in Figure 3) with the capacity to choose cautionary effort x in a way that reduces the bad-joke probability p(x). More contextual detail could bring in other possibilities not considered here such as put-down jokes, 'roast' events, or joke production functions that require dangerous risk taking at the blurry boundary between brilliant delivery of affectionate parody versus offensive insult.

Joke teller chooses cautionary effort x such that bad-joke probability p(x) is decreasing in x

Figure 3 is an extension of Figure 1 (returning from now on to the original assumption of no clairvoyance) in which the joke teller is endowed with a continuously-valued cautionary effort variable, $x \in [0, \infty)$, that effectively reduces the probability of a bad joke. Therefore, p(x) is assumed to be a decreasing function of x. For simplicity, the specific functional form $p(x) = e^{-\alpha x}$ provides intuition for the more general family of decreasing (i.e., controllable) bad-joke risk functions.

Assuming that the unit cost of cautionary effort is measured by parameter k, then Agent 1's expected payoff objective (conditional on telling a joke) can be written as follows:

$$\pi(x) = p(x)b_1 + (1 - p(x))g_1 - kx,$$

which Agent 1 seeks to maximize by choosing x such that telling a joke is better than no joke (i.e., $\pi > 0$) and that $p(x) = e^{-\alpha x}$. The constrained maximization problem just described has a global maximum at $x^* = [\ln(\alpha) + \ln(g_1 - b_1) - \ln(k)]/\alpha$ (in the dense subset of the parameter space of b_1, g_1 and k satisfying the conditions that $x^*>0$ and $\pi(x^*)>0$).

In the models of Figure 1 and Figure 3, we have not considered Agent 1's reasoning about Agent 2's conditions for continuing the relationship or any inferences she makes about Agent 1's intentionality. The interactions so far represented are one-offs unless the payoff parameters are interpreted as depending on both agents' valuations of continuing the relationship in future rounds of interaction. Before proceeding fully toward the fundamental issue of modeling the intentionality of the joke teller, I want to now model Agent 2's decision to either break off the relationship (i.e., not continue) versus continue. Introducing Agent 2's continuation decision turns out to be enough to generate the possibility of the relationship increasing in value following forgiveness in the bad-joke outcome.

Agent 2 chooses "no" or "yes" to continuing the relationship

Figure 4 shows an extension of the basic model in Figure 1 (without a cautionary effort choice x that influences bad-joke probability p), which now includes the possibility of forgiveness and relationship repair in addition to the possibility that Agent 2 chooses to end the relationship by choosing "no." New notation introduced in Figure 4 includes each agents' valuation of the relationship itself, ex payoffs from the joke-telling interaction. The agents' continuation values whenever Agent 2 chooses "yes" to continue are denoted r_1 and r_2, respectively.

Down the event branch in which a bad-joke outcome occurs and Agent 2 decides "yes" to continue nevertheless, then both agents' valuations of their relationship increase to R_1 and R_2, respectively. In keeping with the previous three figures where payoffs represent changes relative to the no-joke state, the continuation value of the relationship does not show up in this representation along the continuation path but instead as a lost continuation value whenever Agent 2 chooses "no."

The loss of the relationship value shows up at nodes where Agent 2 chooses not to continue. At the left-most node, for example, the payoffs are now written as $(-r_1, -r_2)$ if Agent 2 chooses to not continue and (0, 0) if Agent 2 chooses to continue following the no-joke outcome. If, for whatever reason, Agent 2 perceives negative continuation value, then it is rational for her to discontinue because discarding the negative value achieves a positive payoff relative to continuation ($r_2 < 0$ implies $-r_2 > 0$).

At the two nodes following the good-joke branch, the relationship value is assumed not to change and therefore does not show up in the payoffs (g_1, g_2) but is instead deducted from the payoffs as the lost value of continuing the relationship in the payoffs $(g_1–r_1, g_2–r_2)$. Presumably, Agent 2 never has any reason to rationally choose "no" following a good-joke outcome so long as $r_2 > 0$.

The new feature to be analyzed in Figure 4 are the payoffs that follow the bad-joke outcome and Agent 2's decision "yes" (i.e., continuing with the relationship). In this case, in addition to the bad-joke payoffs (b_1, b_2), each agent sees something new about their respective assessments of the value of continuing the relationship. The changes in their relationship values, $R_1 – r_1$ and $R_2 – r_2$, respectively, are added to the bad-joke payoffs. Note here there is a distinct new possibility that the agents that are most well-off (i.e., enjoying the strongest, most durable relationships founded on joint awareness that they are mutually valuable enough to withstand a large range of negative-payoff events and nevertheless retain positive relationship value) are the ones who have endured the bad joke and recovered from it—or simply chosen to continue, thereby revealing far greater implicit continuation values than would otherwise be observable to either player without the failure or mistake.

Result 1: In the interaction represented in Figure 4, if $R_i – r_i > b_i – g_i$, then Agent i is better off after a bad joke is mistakenly told and Agent 2 chooses "yes" to continue the relationship than Agent i would have been without Agent 1 having made the mistake.

It follows from Result 1 that, if the inequality holds for both agents, then the mistake causes a Pareto improvement (i.e., mistakes can unambiguously increase the size of the economic pie by revealing otherwise latent information about the strength of social ties). The possibility that mistakes can strengthen social ties through such a transformatively positive (i.e., relationship-strengthening) act of forgiveness modeled along the bad-joke-"yes" path in Figure 4 brings with it profound implications. Note, too, that instead of forgiveness, the transformative event that occurs can be interpreted as information revelation—that the mistake simply reveals otherwise latent (i.e., unobservable) information about others' subjective valuations of their relationships with us. From this observation, a large set of new mechanisms that map mistakes into aggregate-value-expanding outcomes emerges.

For example, if I fail to deliver on a contractual commitment to a key business partner in a repeated interaction, and that partner expresses understanding and agrees to continue even though I see that my mistake imposed large costs on the partner, then I may be willing to take joint risks with that partner that I would not have otherwise. The reason for the shift in willingness to undertake value-generating risk may be that the process of dealing with my past failure and the hardships it caused both of us transformed the relationship or focused our attention on jointly observing the value of our collaboration. Or it could have simply revealed otherwise unobservable information about my partner's willingness to endure joint losses and remain committed to continuing together, which, in turn, triggers my own willingness to take on new projects where our joint actions expose each other to new risks.

I want to make the case that the interaction in Figure 4 and its scope for generating welfare-enhancing mistakes can be rather broadly interpreted. Telling a bad joke; failing to deliver on a contractual obligation; or being very late in delivering a promised book chapter to an editor whom I respect greatly, whose friendship is dear to me, and whose book project I feel great passion for—these examples are illustrative of smart people's rational mistakes. I will discuss additional examples below. Before considering more examples, I want to begin addressing the as-yet unexamined question of intentionality and why the mechanism of welfare-enhancing mistakes generally breaks down if this mechanism is deliberately exploited.

Model in which Agent 1 can choose to deliberately tell a bad joke

If a prototypical Agent 1 looks at the payoffs in Figure 4 and perceives that the highest possible outcome is indeed the path along which a bad joke occurs and Agent 2 forgives, then could Agent 1 rationally pursue this outcome as his goal? Certainly if Agent 2 knew that Agent 1 were hurting her intentionally in order to coax her into forgiving and revealing her high latent value for continuing with him, then she would likely modify her assessed continuation value of the relationship downward. And if Agent 1 is not sociopathic, then he likely experiences some guilt (denoted $\gamma > 0$, representing Agent 1's psychic cost of deliberately telling the bad joke or otherwise intentionally hurting Agent 2).

To represent intentionality, I introduce the notation $\omega \in \{B, G\}$ to code Agent 1's intention to tell a bad or good joke. Figure 5 depicts, by now, the rather elaborate joke-telling interaction featuring two distinct bad-joke branches which correspond to intentional versus accidental bad jokes. The dotted oval represents Agent 2's uncertainty: When she observes the bad-joke outcome, she does not know whether Agent 1's intention was to tell a bad one or not—intentionally offending and hurting her, or, alternatively, intending to tell a good joke that led accidentally to causing offense or hurt. Because Agent 2's valuations of continuing the relationship now depend on, and vary with, intentionality type ω (through the functions $r(\omega)$ and $R(\omega)$) while holding constant the bad-joke outcome, the model therefore becomes non-consequentialist.

Figure 5 expresses payoffs corresponding to each of the two bad-joke outcomes that differ only in Agent 1's intention ω. But Agent 2 is not clairvoyant and does not know Agent 1's intention (or intentionality type ω) with perfect certainty. In the second dotted oval below the main payoff nodes, Figure 5 also provides Agent 2's expected payoffs (in her state of uncertainty about ω), which depend on Agent 2's probabilistic belief β that Agent 1 is a bad-intention type. The function $R(\omega)$ represents Agent 2's assessment of the value of continuing the relationship with Agent 1 following a bad-joke outcome as a function of Agent 1's type. The function $r(\omega)$ represents Agent 2's assessment of the value of continuing the relationship with Agent 1 along any other dis-continuation or continuation path that does not involve the bad-joke outcome and forgiveness. Note that when Agent 1 is a bad-intention type, both continuation values are assumed to take on very negative and nearly equal payoff values: $R(B) = r(B) << 0$ and $R(B) – r(B) = 0$ whereas $R(G) > r(G) > 0$ and $R(G) – r(G) > 0$ (when Agent 1 is a good-intention type).

In Agent 2's state of uncertainty about Agent 1's type, ω, and having observed the bad-joke outcome, we can compute the difference between Agent 2's expected payoff from choosing "yes" (to continue) minus her expected payoff from choosing "no" (to not continue), which I denote $\Delta_{\text{yes–no}|\text{bad-joke}}$:

$$\Delta_{\text{yes–no}|\text{bad-joke}} = (1 – \beta)R(G) + \beta r(B).$$

Agent 2 (assumed to be an expected payoff maximizer with belief β that measures her subjective probability that 1's type is bad) chooses to continue if and only if $\Delta_{\text{yes–no}|\text{bad-joke}} \geq 0$ (assuming continuation whenever "no" and "yes" have equal expected payoffs) and discontinue otherwise. In other words, Agent 2's continuation decision in the face of being exposed to the bad joke and uncertainty about Agent 1's intentionality type turns on the upwardly-revised relationship value and Agent 1's intentionality type being good, R(G), weighted by 2's belief that 1 is in fact a good type—and then comparing this positive expected continuation value to the negative expected value if 1 were a bad type, r(B), weighted by 2's belief that 1 is in fact a bad type.

Under what circumstances will 1 deliberately cause 2 harm under the expectation that 2 will forgive and continue, thereby yielding a greater player-1payoff than by trying to tell a good joke: $b_1 + R_1 – r_1 – \gamma > p(b_1 + R_1 – r_1) + (1-p)g_1$? Recall that 1 effectively chooses his intentionality type ω. Figure 5 assumes that, if Agent 1 chooses $\omega = B$, then the joke will turn out bad with probability 1. If 1 chooses $\omega = G$, however, then we are back in the non-degenerate probabilistic world where p measures the probability of the bad-joke outcome. A further condition is required if 1 is to believe that 2 will indeed choose to continue. In other words, 1 must believe that $\Delta_{\text{yes–no}|\text{bad-joke}} \geq 0$.

Agent 2's preferences are non-consequentialist, a fact made explicit through the dependence of the functions $R(\omega)$ and $r(\omega)$ on ω. Agent 2's view of her own payoffs is not invariant with respect to w holding the bad-joke outcome fixed. Agent 1's intention—to deliberately tell the bad joke ($\omega = B$) or to at least *try* and tell a good joke ($\omega = G$)—matters quite explicitly to Agent 2.

The next section applies a slightly different interpretation to the payoff schemes in the Figures above to illustrate the general nature of the phenomenon described above in the Figures and Result 1, namely, that individually and collectively welfare-improving mistakes are commonplace and broadly distributed throughout the decision environments people face. The integrity of the mistakes, as distinguished in Figure 5, matters. There are honest mistakes and fraudulent or deliberate ones. Smart agents should, in general, be adept at detecting fraudulent mistakes, although doing so is not necessarily easy in practice.

You're late!

One way I can learn how much you value my work, or my contribution to a joint venture, or our relationship is by observing your willingness to forgive. I sometimes show up late (or as it turns out, deliver work or other outputs much later than originally promised, thereby testing the patience—completely unintentionally—of dear colleagues, people about whom I care deeply and hold in truly great esteem, e.g., by delivering late a chapter for an edited volume on *Behavioral Economics for Smart People*). In response to my lateness, some colleagues may classify me as unreliable and choose not to engage with me on future projects; others may classify me as (once again) unreliable but nevertheless choose to continue engaging with me, effectively revealing that they forgive me for being late, or that they value my contributions highly enough to offset the substantial costs I unintentionally imposed on them with my lateness, regardless of whether forgiveness is formally expressed. In game-theoretic terms, the act of my colleague effectively forgiving my lateness sends an important signal about that colleague's implicit valuation of engaging with me relative to the respective costs that my lateness imposed (again, completely unintentionally on my part).

Why do I emphasize my a priori intention to not be late followed by ex post lateness (i.e., unintended lateness)? Consider re-labeling "bad joke" outcomes in Figures 1-5 with "late." The interactions represented in Figures 4 and 5, for example, can then be reinterpreted as: as long as I am not late, then my payoffs and those of my colleagues correspond to good-joke payoffs, (g_1, g_2), which represents the normal state of affairs based on the productivity of our relationship with no change in trust, no disappointments and no forgiveness. But as soon as I violate my colleague's expectation, the colleague's decision about continuing can then be interpreted as a signal of forgiveness (or otherwise revealing additional mutual value in continuing). Then something new happens.

There is an objective loss to both players: $b_i < g_i$ for $i = 1, 2$. My colleague bears the cost of my lateness equal to $g_2 – b_2$. I pay a cost $g_1 – b_1$ based on embarrassment, loss of reputation for punctuality, and perhaps stress over future opportunities now at risk. But our aggregate payoffs are now mutually recognized as being greater—for both of us—if we continue, thanks to the mutually beneficial interaction of both individuals (assuming the condition in Result 1 holds).

Acknowledging that these costs of lateness can be, and often are, substantial, what then justifies Result 1 and its possibility of greater payoffs—for both of us—corresponding to the action profile of (**late**, **forgive**) with associated payoffs of $(b_1 + R_1 – r_1, b_2 + R_2(G) – r_2(G)) > (g_1, g_2)$?

The answer must be the existence of an offsetting or compensating deepening of the value of our working relationship, where a signal has now been transmitted showing the intention to collaborate cooperatively into the future within a larger-than-expected space of perturbations in the form of missed expectations of various kinds. Another possibility is more direct: the incremental increase in wellbeing that follows from an expression of (relatively) unconditional acceptance.

What have I learned by being seriously late and then receiving implicit forgiveness? I may have learned that my colleague enjoys interacting with me or benefits to a sufficient degree that he or she is willing to incur higher costs than I had perhaps previously realized to keep the working relationship alive. Given the benefit discovered by lateness and subsequent forgiveness, might I then pursue intentional lateness as a mechanism to force colleagues to reveal signals about their willingness to forgive transgressions and maintain working relationships? No. None of this works if lateness (or the bad joke, or any other setback, mistake, disappointment or missed expectation) is intentional.

Suppose I am considering a sequence of lateness decisions, coded as binary for simplicity, with a new person in my life with whom a potentially valuable relationship might unfold. I would like to know how this other person regards me or, more crassly, assesses the potential value of our relationship. In other words, I have positive willingness to pay for a costly-to-fake signal of affection, esteem, or some form of perceived value from continued engagement. The other person would also like such a signal from me. Could I test the other person by deliberately being late, or deliberately telling an offensive joke, for example, to get a live observation of the other person's willingness to forgive?

Intentional mistakes are no longer mistakes, however, and this strategy is unlikely to work. Problems include the high risk of being discovered and my guilt or embarrassment (γ), not to mention the new risk of being discovered, the possibility of an extremely negative payoff that the other person would perceive if I were outed as a perpetrator of intentional lateness, deliberately offensive joke telling, or some equivalently dis-pleasurable breach of the other person's expectations.

Games Against (or in Accordance, Cooperation, Harmony With) Nature

Gigerenzer (2005) discusses physicist Feynman's arguments in favor of violating invariance with respect to logically equivalent re-descriptions of the same problem. Feynman sought out scientifically useful framing effects by which different intuitions about the laws governing a set of variables became more readily apparent using different frames or logically equivalent re-descriptions. He wrote that these logically equivalent re-descriptions are valuable because "psychologically they are different" (quoted in Gigerenzer, 2005, p. 207). In contrast, behavioral economists largely adopt the opposite normative view: that framing effects and other patterns of making different inferences or taking different actions in response to logically identical re-descriptions of the "same" decision problem constitute evidence of irrationality.

Gigerenzer argues that the mind's perceptual system similarly makes smart bets; the intelligence of those bets depends necessarily on making mistakes. For example, in making three-dimensional inferences based on two-dimensional visual input, the mind bets that there is only one source of light that is located above, implying that objects with dark shading below are likely to be "sticking out" toward the observer. From this, Gigerenzer observes that the perceptual system correctly assumes that the world (i.e., its three-dimensional structure) is fundamentally uncertain (i.e., we face the challenge of missing information about three-dimensional structure in our environments) and therefore use associational rules to make reasonable guesses. If instead the mind proceeded as an agnostic Bayesian and waited for irrefutable evidence before logically *deducing* the correct three-dimensional structure, it would be paralyzed. Similarly, if it had access to a veridical copy of all information required to produce an objectively accurate model of all relevant detail in its environment, the mind and its perceptual system would be overwhelmed. The functionality of the simple bivariate-association rule, "objects with dark shading below are sticking out toward me," depends on its partiality and imperfection with respect to veridical descriptive accuracy. A hypothetically perfect (i.e., veridically accurate) perceptual mechanism would still be too little, leaving perceptual holes when facing new or unknown environments (i.e., situations where a quick action based on a snap perceptual bet is required without inputting the vast amount of information that a perfect perceptual mechanism would require). This perfectly veridical perceptual representation of the world would also be overwhelmingly too much, presenting the mind with paralyzingly large volumes of spatial information.

The same goes for memory. Is more better? Not necessarily (e.g., Schooler and Hertwig, 2005, show that forgetting is beneficial in inference tasks). And Gigerenzer (2005) discusses individuals with unusually large recall memory that suffer, as a result of their special pneumonic endowment, with acute inability on tasks requiring abstraction. Perhaps having more recall memory means less practice at efficiently coding the gist of what is taking place, abstracting, and forming equivalence classes in memory.

Are larger consideration sets better than smaller ones? Among the successful entrepreneurs from whom data were collected in Berg (2014b), very small consideration sets with only three potential locations for a high-stakes investment decision were the rule rather than the exception. And larger consideration sets were associated with below-average investment performance. Less was more.

When choosing where to stand to catch a ball, three observations about how professional baseball players do it are noteworthy in that they deviate from how robots would be programmed to do it using a veridical causal model based on initial velocity, wind speed, rotation, etc. Many researchers believe that veridical causal models stand unquestioningly as the gold standard for rational choice. In that view, deviations from how an idealized robot would do it are automatically labeled as mistakes. This view forces the interpretation that the deviations of professional baseball players—who are the best in the world at what they do—are prima facie evidence of irrationality rather than intelligence and high functionality.

If the mind were essentially solving the physics problem of where to stand to catch the ball based on initial velocity, wind speed and rotation, then players who can reliably catch the ball in this way should be able to point to and predict the landing point without actually running to catch the ball. They cannot (see references in Gigerenzer, 2005, and also Berg and Gigerenzer, 2010, on as-if behavioral economics). If players' minds were evolved to approximate the veridical causal mechanism, then they should also run straight to ball's landing spot and do so as fast as they can to leave time for last-minute adjustments.

Instead, they use a gaze heuristic that requires no causally relevant information at all and no precisely optimal angle (but rather allows for a large and forgiving range of angles that function just fine) at which to fix their gaze. The gaze heuristic is a process model: fix the angle of one's gaze to the ball, start running, and maintain the angle. It requires no causally relevant information. And it works.

Bias-variance trade-off

The bias-variance trade-off well known in statistics, machine learning and, more recently, psychology, implies that deliberate bias is, in general, a requirement of virtually any well-performing statistical procedure that fits unknown parameters on a training set and then measures performance in generalization tasks requiring out-of-sample prediction. This trade-off forces the conclusion that insisting on zero bias will lead inexorably to maximal variance which, in any application with a single, finite data set, violates most notions of "well performing."

Lexicographic order of asymptotic consistency over variance reduction in econometrics

Classical econometrics is still taught in many if not most economics Ph.D. programs as if there is a unanimous tacit agreement that the normative criteria of an estimator being unbiased or consistent (asymptotically converging to the correct value with probability 1) is infinitely more important than variance (not to mention performance in out-of-sample prediction). Orthodox econometric pedagogy advances a lexicographic order that ranks biasedness over (i.e., as categorically more important than) variance.

The odd juxtaposition of methodological norms seems noteworthy. Conditional mean functions are specified as flexible but always compensatory functions of the vector of conditioning variables. And utility functions are used that assume preferences cannot be lexicographic. In contrast, in econometrics, however, economists work under the assumption that lexicographic preferences over the characteristics of estimators is reasonable (i.e., unbiasedness and consistency trump any comparisons of variance).

More information not necessarily better even in games against nature

How much information should one pay attention to? And when is it rational to ignore relevant information even when facing no cognitive constraints or costs of conditioning information? Berg and Hoffrage (2008) provide a formal definition of an economic or psychological environment and the matching concept of *ecological rationality*. They demonstrate that there are dense sets of environments in which, because payoffs and probabilities cancel out under the expected payoff operation, a non-redundant predictor or decision cue X that is veridically correlated with future payoffs may nevertheless drop out of optimal action rules, giving rise to the phenomenon of *rational ignoring* environments.

Berg, Biele and Gigerenzer (forthcoming) present data collected from economists that measure both individual-level belief consistency with respect to Bayes Rule and belief accuracy with respect to published point estimates for disease frequencies in the medical literature. Which economists had the most objectively accurate beliefs about prostate cancer risks? It was not the economists whose conditional beliefs were perfectly Bayesian. Formal analysis of the analytic measures of belief consistency and belief accuracy, as well as the empirical data, show that performing well by one of these two distinct criteria does not imply good performance on the other. In many settings the multiple normative criteria that are observable in choice data may be negatively correlated. Perfect time consistency may arise mostly as a result of consistently impatient behavior (so that time consistency and the present value of lifetime wealth or lab earnings are negatively correlated). Perfect conformity with the Savage Axioms may arise primarily as the result of consistently risk-averse choices with far-below average mean earnings. Perfect conformity with transitivity may result primarily as a very clear orientation toward leisure over money, implying that transitive types are, on average, less wealthy, less entrepreneurial and lower earning in lab experiments.

When might the "mistake" of failing to maximize expected utility and satisficing instead lead to social welfare improvements? Berg and Gigerenzer (2007) demonstrate just such an environment. Their model provides a thought experiment about a benevolent social planner: if she were able to choose whether the agents (whom she loves in the sense of wanting them to achieve the greatest possible individual and aggregate payoffs) were expected utility maximizers or satisificers, would the society of maximzers be better off? Berg and Gigerenzer (2007) show that the society of satisificers is unambiguously better off by according to the same social welfare function. The satisificers achieve higher social welfare and require far less paternalistic intervention when compared from the vantage point of a Benthamite social-welfare metric.

Strategic Games Against Self-interested Competitors

Mistakes can make an agent's behavior less predictable and therefore thwart exploitative attacks. Like the television detective Columbo's feigned ineptitude and lack of smarts as a strategy for inducing others to reveal information, agents that adopt decision styles which allow for, and plan on, committing errors can effectively induce adversaries into less cautious play, less aggressive best-response functions, and greater revelation of information. I want to clarify: the errors considered in this essay so far have nothing to do with strategically portraying oneself as stupid. But it is worth including feigned irrationality in this list of examples that illustrate the breadth of mechanisms through which mistakes confer genuine value added. If others are convinced that I am stupid, then I may have more freedom to discover information or trade in markets without others strategizing against me. Inflated or wrong beliefs can make one a stronger negotiating partner. And mistakes lead to discoveries when the environment (e.g., the reward- or payoff-generating process) is changing (Bookstaber and Langsam, 1985).

Markets and Social Systems That Benefit from Logical Inconsistency and Other Alleged Errors

At the species level, sub-optimal individual decisions may be rewarded by what is effectively a species-level portfolio diversification effect. There are some individuals failing to maximize in today's environment, which may seem like a sub-optimal waste. In the event that the payoff environment is buffeted by shocks so that previously optimal behaviors can no longer survive, however, then the currently sub-optimal individuals may come into their own.

Suppose the energy yield from grazing on the north side of the lake is 80 but only 20 on the south side. What is individually rational is, of course, to graze at the north side. At the group level, however, when attacks, pests or poisons can appear on one side or the other, it is adaptive for some individuals to graze on the low-energy-yielding south side. This individual-level mistake averts group-wide cataclysm had all individuals chosen north and an unexpected attack takes place on the north.

Market liquidity itself depends on noise or liquidity traders. Behavioral, belief and preference heterogeneity are primary reasons underlying why trade (i.e., exchange itself) creates economic value. Berg and Lien's (2005) model of Pareto-improving overconfidence in the precision of information possessed by insiders (in beliefs among the uninformed) shows that overconfidence in experts, while sometimes damaging, can generate surprising liquidity benefits in financial markets. These positive externalities in the form of lowered transactions costs more than offset the individual costs of having wrong beliefs. If the otherwise typical payoff functions in their model were alternatively interpreted as representing evolutionary fitness functions, then a striking conclusion emerges: there is no sense in which rational expectations (i.e., objectively accurate subjective probabilistic beliefs) is adaptive; overconfident belief profiles support equilibria that Pareto dominate the rational expectations equilibrium.

Sampling to learn about a changing environment is another benefit of making mistakes. That may explain why experimental subjects who switch their responses (perhaps randomly) to the very same decision tasks at different experimental sessions have been observed to earn more, on average, than consistently impatient and consistently risk-averse individuals do. The consistent types' behavior passes the rationality test according to the norm of internal logical consistency, which is the sole claimant to rationality in rational choice orthodoxy. These consistent individuals earn significantly less, however (Berg, Johnson and Eckel, 2014).

Such contrasts, once again, highlight the multiple normative standards that economists employ, whether tacitly or explicitly (Berg, 2014), in characterizing the rationality of observed choice data. Randomization may confer other surprising benefits, for example, in social systems that offer opportunities for random face-to-face encounters, which Berg, Hoffrage and Abramczuk (2010) show are capable of stabilizing otherwise polarizing social dynamics and preventing Schelling-type dynamics that tend toward absolute segregation.

In public goods games, behavioral economists alternate in their interpretation of what constitutes mistaken behavior. Usually, failing to free ride, as required by the Nash-Equilibrium strategy (under the assumption that all players maximize standard rational choice own-payoff objective functions), is cast as an alleged mistake and serves as one of the main outcome variables that behavioral economists focus on. Kameda et al (2011) report evidence of strikingly intelligent behavior in the nonlinear public goods games they study. The equilibrium in the symmetric public goods games they study requires asymmetric action profiles. Therefore, some means of coordinating or deciding which group member will volunteer to be the sole contributor and agree to be free-ridden upon is required. Rather than widespread pathology, their data reveal a wide range of individual and group intelligence.

Regulation to prevent over-use of a commons is another longstanding question in public economics. The less-is-more principle underlying individual intelligence in Gigerenzer's heuristics reappears, once again, as relevant to regulatory policy across multiple settings. For example, Berg and Kim (2015) show that *permissive* restrictions on overuse of the commons can, counterintuitively, be more effective at mitigating overuse than stricter restrictions would have been, given imperfect enforcement of the regulation.

A similar surprise regarding what looks mistaken through one lens of benefit-cost calculus becoming rational when viewed from another such lens shows up in models of social dynamics that include positive payoffs for coordinating with like types (as well as potentially negative externalities possibly resulting from extreme racial and religious segregation). The Kahneman-inspired normative position of much of behavioral economics condemning human judgment and decision making as generally pathological can be turned on its head once again: rather than widespread pathology as the default normative assessment of behavior that deviates from simple rational choice models and their assumed consistency criteria, there is as yet much intelligence that can be observed in apparently mistaken behavior. Take, for example, the money sacrificed on religious products for which there is an intrinsically equal-value substitute available at substantially lower price. Such behavior can, even without intrinsic benefit, provide socially valuable signaling and coordination functions (Berg and Kim, 2014).

Singular versus plural norms used in defining rationality?

It sounds paradoxical and unbelievable to many behavioral economists, which makes it worthwhile to reiterate: rational choice orthodoxy underlies much of behavioral economics, and the two share a methodological commitment to there being a *single* normative standard of rationality that does not depend on context or domain but instead is decided based solely on internal logical consistency (Berg, 2003; Berg and Gigerenzer, 2010; Berg, 2014). The Kahneman-inspired biases literatures within behavioral economics and the field of judgment and decision making typically focus on deviations from some standard of logical consistency. Behavioral economists working in this vein are generally interested in the observational phenomenon of deviations from such a standard of internal logical consistency.

Rather than question whether this normative standard used to define bias and deviations, the normative validity of the rational choice benchmark remains largely unquestioned among both behavioral economists and proponents of the rational choice orthodoxy. Their shared singular normative standard defines the deviations that comprise the main outcome variables of interest to many behavioral economists. Such standards of rationality based solely on logical consistency include: logical invariance as the rationality standard against which framing effects become interestingly pathological; transitivity as the core component of the rational preference standard against which studies of intransitive and incomplete preference gain traction; Bayes Rule in papers about non-Bayesian beliefs; the logic of set theory in investigations of the conjunction fallacy; and even Nash equilibrium as a benchmark of rationality in hundreds of studies by behavioral economists that report non-Nash play frequencies as the main dependent variable without ever comparing dollar payoffs (or comparisons by other normative metrics) among Nash versus non-Nash subsamples.

In this essay, I am considering the rationality of mistakes and errors of the kinds described above. To do so automatically implies that a newly pluralistic set of normative concepts are required. Ecological rationality is explicitly pluralistic by requiring good-enough (i.e., satisficing levels) of match between a decision procedure and the environment in which it is used. This standard asks that, in a well-specified set of task environments, the decision procedure performs to a functional and pragmatic standard such that, despite and sometimes thanks to making mistakes, the procedure is readily seen as sensible, purposeful and, yes, rational!

In the ecological rationality framework, a particular decision procedure or heuristic is, in itself, neither rational nor irrational. Unlike the rational choice and behavioral economics standard in which a single pair of intransitive choices or violation of logical invariance earns the universal assessment of irrationality, a choice procedure in the ecological rationality framework has performance characteristics that are alternatively rational and irrational depending on the external environment in which it is considered.

It is only once the decision procedure is embedded in a particular environment that Herb Simon's two blades of the ecological rationality scissors (decision procedure and external environment that jointly determine reasonable performance metrics for defining what is good enough to achieve success) can do their work at identifying boundaries that circumscribe the set of task environments in which a particular decision procedure achieves ecological rationality. It appears that any normative framework integrating the possibility of beneficial mistakes as categorized above necessarily implies that pluralistic normative metrics and the adaptive toolkit approach to defining what rationality means are in play.

Which organ in the human body is best?

Does it make any sense to ask which organ in the human body is the best or most valuable one? Using the massively interdependent body as an analogy, the behavioral phenomena of interest to social scientists will generally require multiple normative metrics akin to separately measuring and considering kidney function, liver function, cholesterol, triglycerides, blood glucose levels, etc. Would it make sense to integrate all known organ-specific performance metrics or results from standard blood panels into a single, scalar-valued assessment, perhaps using a label such as Generalized Aggregate Physiological (GAP) score? One is hard pressed to think of any application where such aggregated summaries that compress the body's multiple interdependent systems into a single scalar-valued metric would be more informative or pragmatically useful than the disaggregated components considered as fundamentally multivariate normative outcome.

By analogy, when we ask the normative question using experimental choice data or theoretical models whether an observed set of behavioral patterns could be rationalized as if it were maximizing some scalar-valued objective function with newly exotic preference parameters to more flexibly mop up variation in the data, we are most likely asking a similarly wrong question. The standard analysis of a scalar-valued normative metric asks us to rely on the optimal choice function (i.e., the program that maps exogenous parameters into an endogenous inference or action maximizing the narrowly defined objective). This method leads to the erection of a dug-in methodological phalanx that severely limits behavioral economics to persisting in egregious repetition of what statistician John Tukey called a type-III error: providing the right answer to the wrong question.

In a massively interactive and interdependent biological or social system, the right way to behave depends on context. Rationality norms must be pluralistic and thoughtfully well-matched to a specific (i.e., explicitly delimited) class of decision problems where a particular (i.e., explicitly defined, possibly multivariate) standard of rationality makes sense.

Is Economics The Only Discipline with a Commitment to Mono-Methodological Singularism?

Yes.

Influence by and parallels with the axiomatization program in Mathematics?

The axiomatization program in economics was in part inspired by the axiomatization program of mathematicians such as David Hilbert, Whitehead and Russell, and The Bourbaki Group, which overlaps with the consistency school of normative bounded rationality (Berg, 2014). This axiomatization program profoundly influences (i.e., restricts) economists' normative analysis (i.e., the normative questions that can be asked) in subtle ways that go mostly unnoticed in methodological treatises on the real-world applicability of behavioral economics and bounded rationality. Economic studies of bounded rationality would benefit by noticing the waning trajectory of this axiomatization program in mathematics and, like many in mathematics have, choose instead to pursue applied problems and the informal mathematics described in Backhouse (2008).

I define the axiomatization program in economics as the body of economic theory that seeks a short list of axioms (perhaps minimal in some sense) that exhaustively characterizes the rationality of: preference orderings; sets of observed choices or demanded bundles (the extensive literature on revealed preference typically associated with Paul Samuelson); or orderings on choice sets. This axiomatization program can be narrowed further to investigations that pursue the question of postulating maximally general axioms (i.e., the weakest possible) that can "rationalize" observed choice behaviour. The methodological priority of (topological) generality that characterized much of Hilbert's Program peaked in the latter half of the 20th century. Since then, the dominance of the axiomatic program in mathematics has waned, whereas its methodological force in economics appears to have remained relatively undiminished.

The history of the axiomatization program in economics reflects numerous borrowings and inspirations from mathematicians: David Hilbert, Bertrand Russell, and the Bourbaki Group all sought to rid mathematics of the possibility of inconsistencies. Russell's Paradox provides a primary motivation for early 20th-century mathematicians' program of eliminating inconsistency. That well-known paradox posits a collection of all sets that do not belong to themselves. The contradiction turns on ambiguity in the definition of the aforementioned "collection" enjoying the status of set. By restricting the definition of a set to exclude some otherwise well-defined collections of mathematical objects, Frege, Whitehead and Bertrand, and Fraenkel introduced a new formalism into mathematics to resolve such paradoxes, most often beginning with axiomatization.

There is an even earlier link, however, to the axiomatization program's goal in economics of providing minimal conditions to "rationalize" choice data in Cantor's 1895 theorem. The "characterization" of rationality and the "rationalization" strand of the axiomatization program in economics can be thought of as beginning with a set of axioms and a universe of observable patterns of behavior and then projecting the graph that characterizes all allowable patterns of behavior that satisfy the axioms, which is a strict subset of the larger universe of possible patterns of behavior. This can be backward engineered as follows: Given the observed set of choices or behavior patterns, what axioms must this set of choice data satisfy in order to (i) recover a preference ordering that could have generated the choice data, and (ii) assuming a preference ordering exists for ranking vector-valued bundles or payoff distributions in the case of risky choice, what axioms must the data satisfy for the rankings of multi-dimensional objects to be representable as scalar-valued utility or value scores?

Note that this this rationalization subset of the axiomatization program in economics contains, for example, Tversky and Kahneman's loss-averse cumulative prospect theory, specifically, versions of it that attempted to rationalize the choice data generated in Allais' Paradox (which are interpreted as anomalous with respect to expected utility theory). Rationalizing anomalous choice data is described by Gerd Gigerenzer, Werner Güth and

Reinhardt Selten as a repair program. The goal is to take choice data (from binary choices over pairs of risky gambles in the case of prospect theory as a resolution to Allais' Paradox) that cannot be represented with an expected utility function and then show that those data could have been generated by prospect theory, for some unspecified but theoretically possible parameters that determine the shape of the value-function and the nonlinear function mapping objective probabilities into decision weights. Note that this rationalization project, or repair program, bears some similarity to the fallacy of ranking regression models according to their R-squared. Finding a list of axioms that "can explain" choice data is analogous to a regression model with more right-hand-side variables fitting a dataset better. As econometric textbooks correctly caution, a model that fits the data better may not necessarily make more accurate out-of-sample predictions. Fit can always be made to reach 100% if enough free parameters are added to the model specification, one for each observation in the fitting or training sample.

Cantor proved—more than a century ago—that if a binary relation is linearly ordered, then it is also embeddable as an isomorphism in the real numbers. Technically, this is almost identical to the intellectual work of writing down axioms (i.e., restrictions on the preference ordering) that guarantee representability with utility, expected utility, or prospect-theory value-function scores.

Ragnar Frisch is credited as the first economist to define preferences as binary relations. Contemporary graduate textbooks use very different notation (deleting Frisch's more broad-ranging "choice field" formulation, which distinguishes commodity space from what Frisch referred to as the decision maker's problem space). Frisch played a leading role in the founding of The Econometric Society and the journal *Econometrica*, advocating formalism and math modeling as a primary source of "rigor" needed to put economics on a "scientific" footing (Bjerkholt and Dupont, 2010). Despite his view that mathematizing economics was needed to displace the "verbal" approaches of institutionalists, one is struck by his sophisticated appreciation of the fact that the decisions modeled as constrained utility maximization (exhaustively searching through a feasible set in commodity space) are embedded in a larger problem space that includes problems perhaps not best handled by the techniques of constrained optimization. This notion of a larger "problem space" foreshadows the notion of "environment" used by writers such as Gigerenzer and Vernon Smith in advocating ecological rationality. Frisch's concept of constrained maximization in commodity space as only one decision domain embedded in a larger problem space notably does not appear in most contemporary Ph.D. textbooks, which instead emphasize the flexibility and universality of preference maximization devoid of context specificity.

As is well known to economic methodologists and historians, early representation theorems in utility theory sought to address debates in economics between those who interpreted utility as a potentially measurable psychological metric of hedonic satisfaction and those influenced by logical positivism wanting to remove psychological notions (Bruni and Sugden, 2007). Early representation theorems establishing utility as a purely ordinal concept devoid of cardinal meaning led to representation theorems in expected utility theory, axiomatizations of Bayesian updating as rational belief functions, and, more recently, weaker axiomatizations that can account for (as bounded rational) some well-known anomalies with respect to rational choice theory. It is this last subliterature of economists writing on rationality axioms in behavioral economics and making reference to Herbert Simon's phrase bounded rationality that is relevant to this essay's focus on bounded rationality and smart people's rational mistakes.

It is instructive to recall that the central motivation of Hilbert and Whitehead and Russell's axiomatization program was to formalize mathematics and philosophy with the explicit goal of eliminating inconsistency. Hilbert and Russell undertook this program and advocated that others join them to rid mathematics—and science—of the possibility of generating inconsistent statements, whether those statements be abstract or detailed descriptions of the world.

While Hilbert's move toward formalism profoundly influenced mathematics (and at the same time attracted well-established critics), it eventually waned as new subfields in mathematics applying methods outside the Hilbert Program grew up and gained acceptance as making substantial contributions to mathematics. Applied problem solving, combinatorics, category theory and subfields of mathematics overlapping with computer science achieved influence and prominence while other theorists working in the constructivist and intuitionist traditions similarly produced new knowledge that followed distinct methodological priorities.

The methodological influence of formalism and the axiomatization program in economics followed an arguably equal if not more profound influence in economics (see Backhouse, 1998, regarding formalism in economics versus informal mathematics). One minor parallel between the trajectories of formalism in mathematics and economics was the desire to shed old interpretations (e.g., the interpretation of points, lines and planes in geometry and the psychological or hedonistic interpretation of utility in 19th century economics in favor of utility as a purely ordinal device). Another speculative parallel that can be seen in the restrictions that choice axioms placed on what had been a previously more libertarian view of consumer sovereignty is to see them echoing the restrictions that Frege, Whitehead, Russell, Fraenkel and Hilbert applied to the definition of a set (in order to avoid paradoxes such as Russell's Paradox).

Beyond these similarities, however, the differences in the historical trajectories of axiomatization programs in mathematics and that of economics are many. Formalism in economics (until very recently) did not have a long struggle with concepts such as: the definition of a set as its core methodological problem; syntactical formalism; the incompleteness theorems of Gödel; and many others. The mathematical issues in the development of economists' formalism were, by and large, far simpler mathematically, and focused on applying topological formalisms already established in mathematics to preferences and representations of preferences. In the axiomatization program in economics, the role of interpretation and motivation of axioms were the primary objects of notable theoretical economists' writing. Critiques and crises over the roles of an axiomatization program (and the "interpretation-free" view of mathematics as a content-free set of primitives and a formulaic set of statements based on definitions of operations juxtaposed or concatenated to generate all permutations allowed by the axioms) did not surface or echo in economics, at least in obvious ways.

These differences, however, serve to cast into sharp relief the one over-riding similarity between the axiomatization programs of math and economics: internal logical consistency as the pre-eminent normative value.

Behavioral economics is normal science portending no paradigm shift in normative analysis

Some argue that behavioral economics should be interpreted as a paradigm shift or otherwise momentous contestation in reaction to the axiomatization program in economics. Behavioral economists' work could, if such an interpretation were granted, be seen as echoing earlier methodological shifts in mathematics following the rise and decline of the Bourbaki Group's influence in mathematics in the 20th century. I want to argue against this methodological view of behavioural economics as a paradigm shift and instead demonstrate that its over-riding normative value remains firmly rooted in the axiomatization program's normative view, namely, that the central concern is, and should be, internal logical consistency.

Those who see behavioral economics and modelers of bounded rationality acting as an ensemble to "expand" or "loosen" the methodological strictures of rational choice theory miss a crucial difference in the normative views of the consistency and ecological rationality schools as I have defined them in earlier work (Berg, 2014). Behavioral economists in the consistency school propose radically narrow normative definitions of rationality and use the label "bounded rationality" (in a manner that would seem to contradict Herbert Simon's normative view). The result is to harden the methodological commitment to internal consistency as the sole criterion that economists are expected to use in characterizing what it means to make rational decisions—and in prescriptive policy proposals that paternalistically intervene, aiming to induce people's private actions to more closely conform to axiomatic models of rationality.

Backhouse (1998) reminds us that axiomatization, mathematicization and formalization are distinct. Gigerenzer and Selten's (2001) ecological rationality program provides a clear example of normative decision analysis that draws on quantitative data to produce theories that can be expressed in the language of mathematics, yet have nothing to do with axiomatization. Backhouse notices (as many other writers on mathematics and philosophy, and the history of mathematics have) that mathematics itself can be either formal or informal.

In the development of proofs of Euler's theorem, for example, which relates the numbers of vertices, faces and edges of a polyhedron, Backhouse (1998) describes different authors' proofs as somewhere "between formalism and irrationalism. . . . There is more to mathematics than driving the properties of formal systems."

The implication would seem to be that applied economics, welfare economics and prescriptive policy analysis cannot be entirely about deductive logic (Berg, 2007). Indeed, the proper role of deductive logic led to animated and productive debates about mathematical methodology and philosophy regarding the Hilbert Program among constructivists, intuitionists (including Hilbert's students Brouwer and Weyl), subsequent work in proof theory, category theory and those inspired by Turing on computability.

Given these prolific bodies of work by mathematicians that raised questions about consistency as the core methodological concern in mathematics, it would seem wrong for economists to draw the lesson from mathematics—in the name of "providing rigor" or "putting economics on a more scientific basis"—to insist on applying consistency alone as the ultimate methodological value.

What are we to make of the long tradition among neoclassical economists—and now behavioral economists—who seem to follow Hilbert's singular normative premise in pursuit of logical consistency? I think we can note the positions of economists like Debreu and Binmore as playing a role similar to Hilbert's role in mathematics. Their staunch position in favor of consistency as a singular methodological and normative-prescriptive value is simply one among multiple, competing normative claims within economics. Heterogeneity of methodological priorities is a positive symptom of productive scientific investigation. In light of the productivity generated by those who raised questions and took positions against Hilbert's consistency program in mathematics, however, economists might also notice that competing normative claims are likely to play a similarly productive role in economics. Such methodological debate is no small side issue but rather a substantial object of core investigation in normative economics.

Concluding Remarks

Casual empiricism and the theoretical economics, biological sciences and biostatistics literatures provide a rich collection of source material from which one finds a broad range of mechanisms by which smart people make rational mistakes. Additionally, economies that generate value added and nurture richly multi-dimensional measures of wellbeing generate numerous opportunities by which aggregate performance is enhanced thanks to systematic deviations from standard rationality criteria based solely on internal logical consistency. I have provided examples that hopefully give a sense of the technical, substantive and historical range of context-specific mechanisms in which alternative normative criteria that allow for welfare-enhancing deviations from logically consistent axiomatic rationality can be given even-minded consideration. May further study of this important phenomenon bloom forth and melt away the methodological strictures unnecessarily limiting behavioral economists' evaluations of rationality.

References

Backhouse, R. E. (1998). If mathematics is informal, then perhaps we should accept that economics must be informal too. The Economic Journal, 108(451), 1848-1858.

Berg, N. (2014a) The consistency and ecological rationality schools of normative economics: Singular versus plural metrics for assessing bounded rationality, Journal of Economic Methodology 21(4), 375-395.

Berg, N. (2014b), Success from satisficing and imitation: Entrepreneurs' location choice and implications of heuristics for local economic development, Journal of Business Research 67(8), 1700-1709.

Berg, N. (2007) Behavioural economics, business decision making and applied policy analysis, Global Business and Economics Review 9 (2/3), 123-125.

Berg, N. (2003) Normative behavioral economics, Journal of Socio-Economics 32, 411-427.

Berg, N., Biele, G. and Gigerenzer, G. (forthcoming), Consistent Bayesians Are No More Accurate Than Non-Bayesians: Economists Surveyed About PSA. Review of Behavioral Economics (ROBE).

Berg, N. and Gigerenzer, G. (2010) As-if behavioral economics: Neoclassical economics in disguise?, History of Economic Ideas 18(1), 133-166.

Berg, N. and Gigerenzer, G. (2007) Psychology implies paternalism?: Bounded rationality may reduce the rationale to regulate risk-taking, Social Choice and Welfare 28(2), 337-359.

Berg, N. and Gigerenzer, G. (2006) Peacemaking among inconsistent rationalities?, In Engel, C. and Daston, L. (Eds.), Is There Value in Inconsistency?, Baden-Baden: Nomos, pp. 421-433.

Berg, N. and Hoffrage, U. (2008) Rational ignoring with unbounded cognitive capacity, Journal of Economic Psychology 29, 792-809.

Berg, N., Hoffrage, U., and Abramczuk, K. (2010) Fast Acceptance by Common Experience: FACE-recognition in Schelling's model of neighborhood segregation, Judgment and Decision Making 5(5), 391-410.

Berg, N. and Kim, J.Y. (2015), Quantity restrictions with imperfect enforcement in an over-used commons: Permissive regulation to reduce over-use?, Journal of Institutional and Theoretical Economics 171(2), 308-329.

Berg, N. and Kim, J.Y. (2014), Prohibition of Riba and Gharar: A signaling and screening explanation?, Journal of Economic Behavior and Organization 103, 146-159.

Berg, N. and Lien, D. (2005) Does society benefit from investor overconfidence in the ability of financial market experts?, Journal of Economic Behavior and Organization 58, 95-116.

Bjerkholt, O. and Dupont, A. (2010). Ragnar Frisch's Conception of Econometrics. History of Political Economy, 42(1), 21-73.

Bookstaber, R. and Langsam, J., (1985), On the optimality of coarse behavior rules, Journal of Theoretical Biology 116, 161-193.

Bruni, L., & Sugden, R. (2007). The road not taken: How psychology was removed from economics, and how it might be brought back. The Economic Journal, 117(516), 146-173.

Gigerenzer, G. (2005). I think therefore I err. Social Research, 72, 195–218.

Kameda, T., Tsukasaki, T., Hastie, R., and Berg, N. (2011). Democracy under uncertainty: The wisdom of crowds and the free-rider problem in group decision making, Psychological Review 118, 76-96.

Schooler, L. J., and Hertwig, R. (2005). How forgetting Aids heuristic inference. Psychological Review 112, 610-628.

Figure 1: Small-world event tree with no possibility of intentionally telling a bad or offensive joke, with bad jokes occurring as an act of nature with probability p, 0 < p < 1

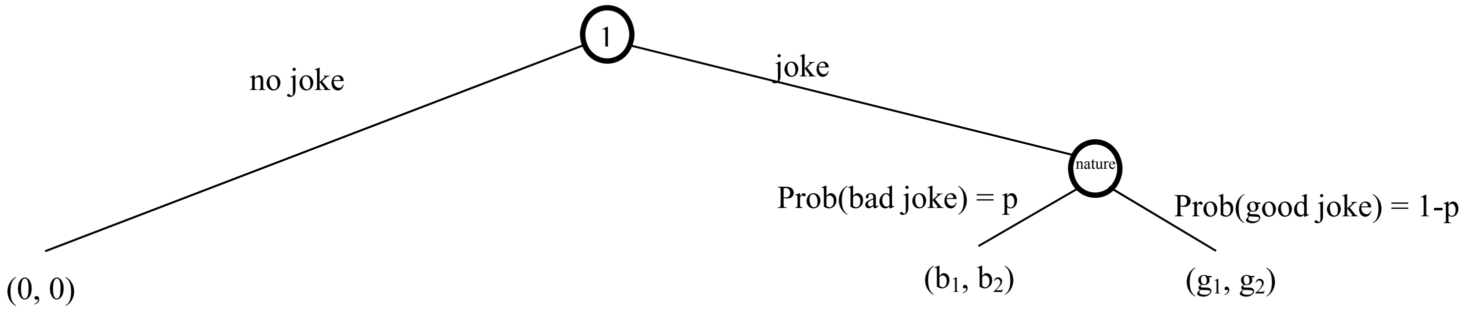


Figure 2: Nature moves first (deciding whether Agent 1's joke will turn out to be good or bad) or, equivalently, Agent 1 is clairvoyant

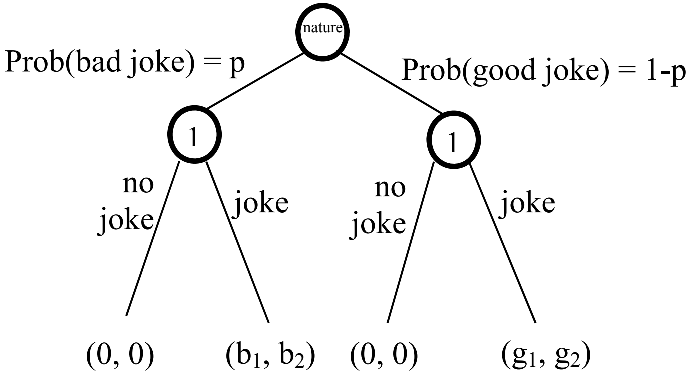


Figure 3: Joke teller chooses a continuously-valued cautionary effort variable, $x \in [0, \infty)$, such that the bad-joke probability is $p(x) = e^{-\alpha x}$

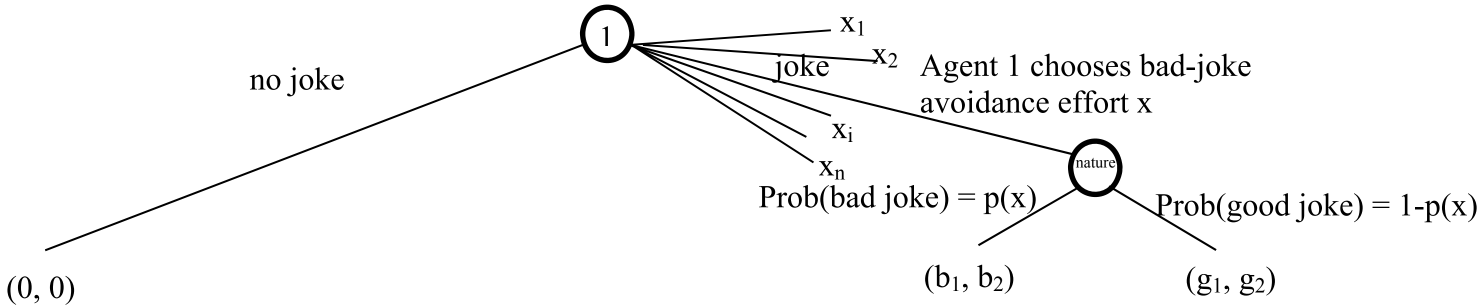


Figure 4: Same as Figure 1, but Agent 2 now chooses whether to continue the relationship ("no" not continue or "yes" continue), depending on whether her relationship valuation r_2 remains positive, with forgiveness of bad jokes having the effect of increasing both agents' relationship value to $R_1 > r_1$ and $R_2 > r_2$, respectively

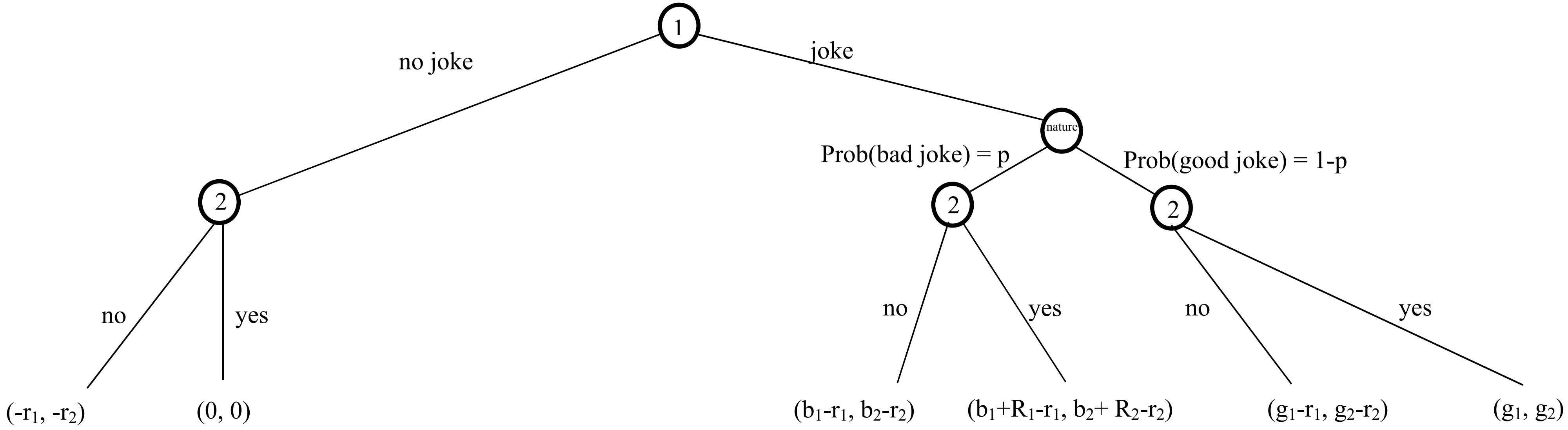
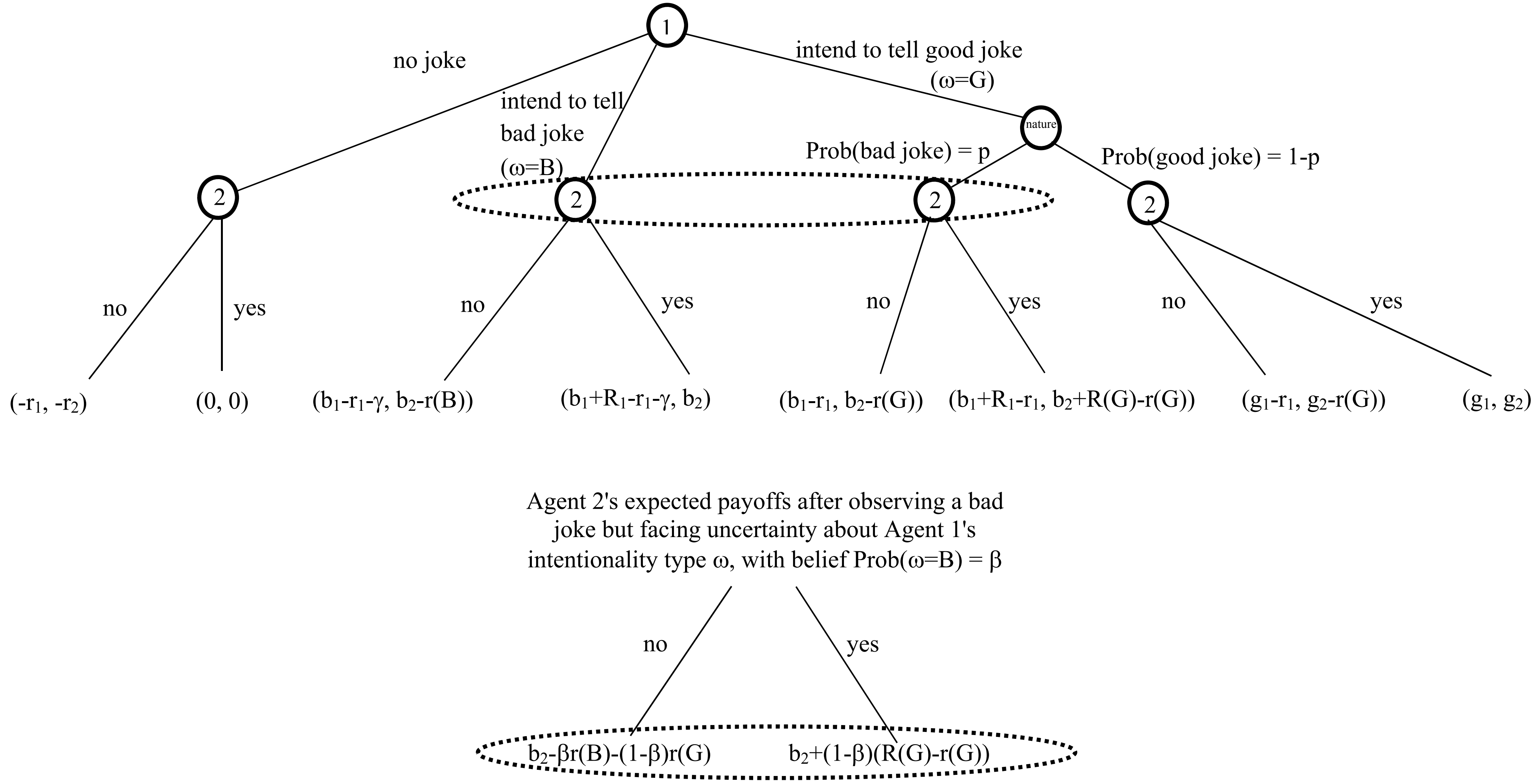


Figure 5: Agent 1 can deliberately tell a bad joke


1
no joke
intend to tell bad joke (ω=B)
intend to tell good joke (ω=G)
nature
Prob(bad joke) = p
Prob(good joke) = 1-p
2
2
2
2
no
yes
no
yes
no
yes
no
yes
$(-r_1, -r_2)$
$(0, 0)$
$(b_1-r_1-\gamma, b_2-r(B))$
$(b_1+R_1-r_1-\gamma, b_2)$
$(b_1-r_1, b_2-r(G))$
$(b_1+R_1-r_1, b_2+R(G)-r(G))$
$(g_1-r_1, g_2-r(G))$
(g_1, g_2)
Agent 2's expected payoffs after observing a bad joke but facing uncertainty about Agent 1's intentionality type ω, with belief Prob(ω=B) = β
no
yes
$b_2-\beta r(B)-(1-\beta)r(G)$
$b_2+(1-\beta)(R(G)-r(G))$