Rev Austrian Econ (2013) 26:329–345
DOI 10.1007/s11138-012-0193-5

Can probability theory deal with entrepreneurship?

Vlad Tarko

Published online: 30 October 2012

Abstract The Austrian theory of entrepreneurship emphasizes the importance of epistemic heterogeneity and the unlistability of the set of all possibilities. A similar concern with what has been called "the art of choosing the space of possibilities" is an important part of Bayesian model selection. Both Austrian and Bayesian authors view the common knowledge assumption as an unrealistic and unnecessary restriction. This coincidence of concerns leads to a joint theory of entrepreneurship. Three important benefits result from this merger: (1) the ability to use Itti & Baldi's Bayesian theory of surprise to empirically measure radical surprise and improve the Betrand competition model as a consequence, (2) dealing with the unlistability problem, and (3) better understanding why the emergence of common knowledge is always the outcome of a social process rather than an inherent consequence of "rationality".

Keywords Kiznerian entrepreneurship · Bayesian surprise · Unknown unknowns · Radical uncertainty · Bertrand competition

JEL Classification D01 · D83 · D84 · C11

1 Introduction

On average, about 25 % of new firms don't survive the first year, less than half survive past the third year, and less than 30% survive past 10 years (Headd 2003; Knaup 2005; Shane 2008: p. 99). Similarly, less than half of self-employed persons remain self-employed past the 6 years mark, with a long term success rate of about 40% (Shane 2008: p. 99). Nonetheless, one in four entrepreneurs who close their business try again another venture (Schutjens and Stam 2006). Given this rather high

V. Tarko (✉)
Economics Department, Mercatus Center, George Mason University, 4400 University Drive, 3G4, Fairfax, VA 22030, USA
e-mail: vtarko@gmu.edu

rate of closures and perseverance in what, at least at first glance, seems like error, we are lead to ask: Are people starting new businesses overconfident and deluding themselves about their chances of success (Wu and Knott 2006)? In other words, does the entrepreneurial engine of our economy depend in some fundamental way on the irrational urge of some people?

There are three theoretical positions with respect to this question, three ways in which one can attempt to contest the apparent irrationality of entrepreneurship, either placing it in a probabilistic framework or by denying that probability theory can legitimately be applied to the issue of entrepreneurship. Entrepreneurship can be addressed probabilistically either by using a rational expectations (objective probabilities) approach or a Bayesian (subjective probabilities) approach. By contrast, following Knight's distinction between risk and uncertainty, entrepreneurship remains in the realm of the unmeasurable. These three approaches lead to very different strategies of trying to assess the rationality of entrepreneurship. The present paper explores mainly the Bayesian approach, focusing on the differences in background assumptions that entrepreneurs make, and shows how the Austrian theory of entrepreneurship (usually following Knight) can be enriched.

The paper proceeds in the following manner. Section 2 highlights the main differences between the three perspectives on entrepreneurial rationality. Section 3 briefly explains why the Knightian approach is outdated (and has been so for a long time) and restates the risk-uncertainty distinction from the Bayesian perspective. Section 4 distinguishes neoclassical surprise from Austrian ("radical") surprise and shows how to use the existing "Bayesian theory of surprise" Itti and Baldi's (2005, 2006) to measure both of them. It is thus shown that, contrary to critics who have claimed that Austrian surprise is a hopelessly vague concept (Caplan 1999, 2001), it is in fact in no way more mysterious or hard to measure than the more familiar neoclassical surprise (low probability events happening). The ability to measure radical surprise leads to an improvement in the Bertrand competition model. Section 5 shows that the unlistability problem emphasized by Kirzner (1997) or O'Driscoll and Rizzo (1985), as an objection to the use of Bayesian theory, is actually far less serious or untractable. Section 6 explores one of the most important consequences of the present theory: Bayesian theory is not only able to model epistemic heterogeneity, but it also makes clear why the emergence of common knowledge is not automatic, happening merely as an inherent consequence of "rationality", but requires instead the working of a social process of the type proposed by Boettke and Coyne (2009).

2 Three approaches to the question of entrepreneurial rationality

From a neoclassical rational expectations point of view, a rational entrepreneur would take the observed frequency of successful new firms as an estimate of one's own success probability. However, the high *closure* rate is not necessarily evidence of *failure* (Headd 2003). On one hand, (a) a business may be closed because investment has reached its optimum end time, i.e. the diminishing returns point beyond which expected profits are lower than the interest rate (Hirshleifer 1970: pp. 81-7), and, on the other hand, (b) entrepreneurship can also be understood as a consumption good rather than as an investment good, with entrepreneurs drawing significant utility from the activity itself, regardless of whether it is profitable or not (Sarasvathy 2001, 2003; Sarasvathy and Dew 2007). Form this neoclassical perspective, whatever heterogeneity exists among entrepreneurs, it is entirely due to *preferences heterogeneity* with respect to the consumption good aspect of entrepreneurship.

The classic Austrian view, based on the uncertainty-risk distinction proposed by Knight (1921), is diametrically opposed to the neoclassical perspective (Kirzner 1997). The most important objection Austrians have to rational expectations regards the assumption of common knowledge. They note that entrepreneurship essentially depends on *epistemic heterogeneity*, i.e. on the fact that various people, due to their different ways of understanding and interpreting the world around them, see different opportunities for profit (Kirzner 1973, 1976, 1980, 1997; Lachmann 1976; O'Driscoll and Rizzo 1985; Boettke 1998; Klein 2012). Their position is to deny that one can have a meaningful statistical estimate of the risk of starting a new business. Due to the high heterogeneity of the contexts in which any new business starts, the uncertainty involved in such a decision is *unquantifiable* and, thus, entrepreneurship lies outside the scope of cost-benefit analysis—it is *neither* rational nor irrational. As Knight put it (1921: III.VII.47), "[b]usiness decisions … deal with situations which are far too unique, generally speaking, for any sort of statistical tabulation to have any value for guidance. The conception of an objectively measurable probability or chance is simply inapplicable."

Finally, the Bayesian perspective, advocated by this paper, is situated in-between the other two extremes. On one hand, it agrees with the Austrians about the importance of epistemic heterogeneity and imagination for understanding entrepreneurship. On the other hand, it agrees with the neoclassical perspective that the probability of success is quantifiable, and thus that cost-benefit analysis is possible even in the case of a decision to start a new business in a novel context or a decision to introduce a new product or service on the market. However, the Bayesian perspective is not a *compromise* between the other two. Instead, it stems from rejecting a common assumption about the objectivity of probability estimations that the rational expectations approach and the Knightian-Austrian approach both share. From the Bayesian perspective, a probability distribution is not an objective feature (or property) of the outside reality which one tries to "measure"; probability distributions are quantitative descriptions of the *state of knowledge* that one has—i.e. they describe the existing epistemic *relation* between an observer and his or her environment, rather than just the objective state of affairs in which the observer is situated (Jaynes 2003: chapters 1 and 2; Samuelson 2004: section 3).

What this means in technical terms is that *all probabilities are conditional probabilities*. It is this relativity of all probability estimations to specified conditions (which can be either hypothetical assumptions or empirical information) that creates the analytical ability to deal with epistemic heterogeneity, as different agents A and B make different probability estimates of the same event x due to their different background assumptions: $p(x|A) \neq p(x|B)$. From this perspective, entrepreneurial action is a combination of two tasks: (1) the task of selecting one's assumptions about the world (the imagination part), and (2) the standard constrained maximization process of cost-benefit analysis (the mechanical part). The reason why different people are not as good at detecting opportunities (of being "alert") is that they have different background assumptions, which lead them to different probability estimations of the likelihood of success of a particular type of action. Thus, understanding differences in alertness amounts to nothing more but understanding that people make probability estimations conditional on different assumptions.

3 The Bayesian risk-uncertainty distinction

There is an important sense in which the Bayesian approach to probability (Jaynes 1988, 2003) undermines the Knightian original distinction between risk and uncertainty (Knight 1921). From a Bayesian point of view, Knight had understood risk—the measurable side of uncertainty—in an overly restrictive manner. This was the result of him accepting the logical positivist version of probability theory which contended that the only "legitimate" way of measuring probabilities is by frequency considerations in repeated experiments under homogenous conditions (Carnap and, more recently, Suppes 1986, are notable exceptions among the logical positivists). Knight adopted this view as the very foundation for his distinction between risk and uncertainty (1921: III.VII.23; III.VII.46-8). This view implies that whenever one *cannot* set up a series of repeated experiments under similar conditions, one cannot define and measure probability, and, thus, one finds oneself in a situation of genuinely unmeasurable uncertainty. The core of the classical Misesian critique of the use of probability theory in economics, as well as of Hayek's critique of the use of probability theory in the social sciences in general (Mises 1996 [1949]: chapter 6; 1957 [2007]; Hayek 1952) is already present in a nutshell in Knight's account, thanks to his reference to the "homogeneity in our groups of instances".

These Knightian-inspired critiques of the use of probability in economics and the social sciences are seriously weakened once one adopts the Bayesian perspective of Laplace (1840), Keynes (1921), Jeffreys (1961) and Jaynes (1988, 2003) (see Langlois 1982a, b; Caplan 1999; and Crovelli 2009, for details, quotes and a critical assessment). According to this alternative perspective, there are numerous *other ways* in which one can measure probability, apart from repeated experiments under homogenous conditions. These involve everything from the principle of indifference and symmetry considerations (which allow one to determine an event's probability even before making a single sampling experiment) to entropy maximization, the most general method developed so far (see Kass and Wasserman (1996) for a critical review of available methods).

Given this background about the way in which the distinction between risk and uncertainty has been originally made, and the realization that the original concept of risk itself has been defined based on an overly-restrictive method of assigning probabilities, there is a legitimate suspicion that perhaps the risk-uncertainty distinction itself is mistaken (Caplan 1999, 2001). In other words, perhaps that, if we take into consideration the more general Bayesian perspective, *everything* that was once thought to be unmeasurable will turn out to be measurable. One should take "guesstimation" methods much more seriously and recognize the fact that one's ability to measure something often depends only on how ingenious one manages to be (Hubbard 2010): apparently "intangible" variables often turn out to be quite measurable if one corners the matter from enough empirically available directions.

This conclusion would be premature. Far from undermining the distinction between risk and uncertainty, the basic Bayesian insight that all probabilities are conditional probabilities leads unavoidably to it. The idea that all probabilities are conditional means that there is no such thing as $p(x)$ but only $p(x|X)$, where X is the background information used to compute the numeric value of the probability of x—the set of hypotheses on which we are relying when we assign a probability estimation to the truth of a statement. At a minimum, X is the assumption about the set of possible values that x can take, but it may include much more complex assumptions as well. As Keynes (1921: p. 3–4) explained:

> All propositions are true or false, but the knowledge we have of them depends on our circumstances; and while it is often convenient to speak of propositions as certain or probable, this expresses strictly a relationship in which they stand to a *corpus* of knowledge, actual or hypothetical, and not a characteristic of the propositions in themselves. A proposition is capable at the same time of varying degrees of this relationship, depending upon the knowledge to which it is related, so that it is without significance to call a proposition probable unless we specify the knowledge to which we are relating it.

In other words, the "true probability", to which Knight (1921) constantly refers, simply does not exist (Phillips 1970). Similarly, the reason why a reluctant Bayesian like Lucas (1977) always puts "true" probability in scare quotes is that he knows that non-conditional probability is pure myth. Probability *always* describes one's knowledge about the world and not a purely objective property of the world (a "propensity" as Popper called it). Interestingly, this is so even in the simplest possible case of probability estimations from repeated experiments under homogenous conditions. Even these probabilities are conditional on one's hypothesis about which set of possibilities one is exploring by means of the repeated experiment (Jaynes 2003: chapter 3; Jensen and Nielsen 2007: chapter 6; annex of the present paper).

Once we take into consideration the conditional nature of all probabilities, the distinction between risk and uncertainty can be restated without relying in any way on the contentious requirement of repeated experiments under homogenous conditions. The risk of x is the value of its probability computed for a *given* set X of assumptions (hypotheses):

$$risk(x) = p(x|X)$$

The uncertainty of x refers to the possibility that the background information X, on which we are relying in order to compute risk, might not actually be correct. In other words, we must consider all possible versions (be they infinitely many) of the background information:

$$uncertainty(x) = \{p(x|X_1), p(x|X_2), \ldots\}$$

This approach seems to favor a radically relativist perspective. If we're basing our predictions on some estimate of risk, dependent on the hypotheses set X, we may turn out to be completely off-track if X is actually false. What is missing is the recognition that, after all, *there exists a Bayesian model selection method* (Bretthorst 1996). The set *uncertainty*(x) is thus somewhat deceptive because its components, the probability distributions $p(x|X_i)$, don't have an equal standing—some are more plausible than others. However, this cannot solve the problem entirely. The question arises: They are more plausible *conditional on what assumptions*? We can nonetheless push the issue one step further, considering the probability of each hypotheses set X_i:

$$uncertainty(x|B) = \sum_i p(x|X_i, B)p(X_i|B)$$

This obviously does not fully reduce the concept of uncertainty to risk, because it still depends on the meta-hypotheses B about how to construct the set of all possible hypotheses X_i. What this shows is that *it is impossible to completely get rid of uncertainty*. The reader might notice the similarity between this approach and O'Driscoll & Rizzo's "The Nature and Process of Learning" (1985: pp. 37-8), in which X_i correspond to different "frameworks", and the overall background B can be created from the X_is by means of a family resemblance approach.

To sum up, risk refers to probability estimations based on a *given* set of hypotheses, while uncertainty refers to the possibility that one's set of hypotheses may contain errors or important omissions. Importantly, the connection between this distinction and entrepreneurship is preserved: entrepreneurship involves differences between peoples' hypothesis spaces, i.e. stems from differences between their representations of the world (including the social world). Kirznerian alertness exists because, thanks to one's different background assumptions, one assigns a higher probability to some opportunity for profit as compared to other people who use some other hypothesis spaces. Similarly, Kirznerian error reflects the fact that one's subjective probability estimations (i.e. risk estimations) may lead to completely mistaken predictions as a result of the fact that they are computed on the basis of some mistaken assumptions. To put it differently, one's ability to identify "opportunities", to be "alert", is a consequence of one's representation of the world. Uncertainty exists because one may misidentify the set of possibilities (i.e. fail to identify one's opportunities), and one's models may be flawed.

4 The mathematical measure of surprise

The above perspective on the risk-uncertainty distinction and the restatement of Kirznerian entrepreneurship allows us to create a rigorous measure of surprise. The Austrian concept of "radical" or "sheer" uncertainty has been accused of being hopelessly vague (Caplan 1999, 2001). However, the Bayesian perspective proposed in the previous section opens the door to measuring this radical uncertainty in the same way, and with the same ease, as the non-radical uncertainty and surprise.

Neoclassical surprise refers to the fact that low probability events sometimes happen and, as Bryan Caplan (1999) put it, this is "inherently surprising". Indeed, when one wins the lottery, one is surprised, although the probability of that happening was (roughly) known. In Bayesian theory, this corresponds to the well-known Itti and Baldi's (2005, 2006) theory of surprise, according to which the surprise level of a new piece of information I (with respect to some matter of interest x) is determined by the difference between the prior quantity of information, $\log_2 p(x|X)$, and the posterior quantity of information, $\log_2 p(x|I, X)$.[1] The unit of measure for surprise is called, aptly enough, a "wow", with one wow corresponding to a two-fold increase from the prior to the posterior probability.

Austrian or "radical" surprise, by contrast, refers to the fact that one may discover that the probability distribution itself has been improperly calculated—a conclusion which emerges from the bad consequences of using that probability distribution. In other words, it may turn out that X is not entirely correct and has to be replaced by Y, with the probability distributions thus changing to $p(x|Y)$ and $p(x|I, Y)$. We can measure the Austrian surprise in wows as well, but this time with respect to the difference between $\log_2 p(x|X)$ and $\log_2 p(x|Y)$, or, equivalently, between $\log_2 p(x|I, X)$ and $\log_2 p(x|I, Y)$.

The Austrian surprise (at least in this Bayesian interpretation) is thus in no way more mysterious or less measurable than neoclassical surprise. The only difference is that the change of the probability distribution is caused by a different kind of change in the underlining conditions (assumptions and/or data). Note also that there is a certain conceptual overlap: if Y is nothing more than the conjunction between X and the new information I ($Y = X \,\&\, I$), then the surprise can be interpreted, equivalently, as either neoclassical surprise or Austrian surprise. This may be a source of some confusion, leading some to conclude that Austrian surprise is actually nothing but neoclassical surprise in disguise. However, this is not necessarily so, as the transition from X to Y can be much more complicated, as it has been pointed out for example by O'Drisscoll and Rizzo (1985: pp. 37-8). In general, the two kinds of surprise are genuinely different.

Finally, the *overall* level of surprise is measured as the change from $\log_2 p(x|X)$ to $\log_2 p(x|I, Y)$, and it is simply the sum of the two kinds of surprises computed separately. Interestingly, the fact that Austrian surprise is also measurable shows us that *an entrepreneur can estimate the extent of the miscalculation by the other market participants* (who are yet unaware of the innovation that she is about to introduce). This has two economic consequences, one philosophical (or methodological) and one economical.

The philosophical consequence is that one can now explain away *within the rational choice framework* various puzzling and apparently irrational decisions of entrepreneurs. This is harder or impossible to do in the rational expectations model or in any model that works under the common knowledge assumption. Take for example Mark Zuckerberg's decision to refuse to sell his business even for sums of money that seem extremely large to outside (and perhaps less informed) observers. One interpretation of such refusal is that he is absurdly vain. However, another interpretation is that his refusal to sell is a signal that he considers Facebook to be far more disruptive than the potential buyers have estimated it to be (this is how some commentators have indeed interpreted his refusal). This rational choice explanation thus repeats at a deeper level the same line of thinking that lies behind the no-trade theorem (Milgrom and Stokey 1982; Samuelson 2004), without however leading to the same no-trade conclusion, because the assumption that all agents share the same priors is no longer part of the story.

[1] The logarithm is taken in base 2 in order to measure the quantity of information in bits.

On a more practical level, this method of measuring radical surprise creates a better tool for planning within the firm or, from an economic theory perspective, it adds another element to the mathematical theory of competition. Consider a Bertrand model of competition in which firms are price setters instead of price takers. When a firm innovates and plans to introduce a new product or service on the market, it does three things: [2] (1) It makes a more or less wild guess about the market demand for the new possible product or service x. (2) It determines its own technical efficiency with respect to producing x, i.e. it estimates the probable marginal cost MC_x of producing x. (3) It decides how much to produce, Q_x, and, based on its estimation of consumer demand, at what price to sell, P_x, depending on (a) its guess about the ease with which the product x can be copied by competitors and (b) its discount rate of the future. The firm sells x at some monopoly price, $P_x > MC_x$, but the monopoly price attracts competitors to the market in the standard Hayekian fashion. As the competitors start selling similar products, the profits reach zero after a time period T. The firm chooses $Q_x\ (t)$ and $P_x\ (t)$ such that it maximizes its entire future stream of expected profits, $\pi(Q_x(t), P_x(t)) = Q_x(t)[P_x(t) - MC_x(t)]$.

The procedure at stage (3) crucially depends on the estimation of the ease with which competitors will be able to enter the newly created market. For the other two parts there already exist various estimation methods, e.g. the probable consumer demand can be estimated from the results of focus groups. The third part is the most difficult one. My suggestion is that one can take the measure of radical surprise described above as a starting point for estimating the ease of market entry on the part of competitors. In other words, the innovating firm is in a unique position on the market to know exactly what background assumption needs to be altered in order to get to their new product. As such, it can estimate how many "wows" the new product will generate, and it can assume that (at least roughly) the more surprising their product is, the more difficult it would be for the other firms to adapt and compete. One can basically see this as a quantitative version of the theory of creative destruction.

Thus, the combination between the Austrian concept of radical surprise and the Bayesian method of measuring surprise allows us to add a missing piece to the mathematical theory of Bernard competition.

5 Is there an unlistability problem?

It is worth noting that some of the modern Austrians have acknowledged the problem with the original Knightian distinction between uncertainty and risk (O'Driscoll and Rizzo 1985: p. 75): "The critical contrast is therefore not … between measurable and unmeasurable uncertainty (Knight 1921), or even between subjective and objective interpretations of probability … ." Following Kirzner, they have proposed instead a different basis for the distinction (p. 71):

[2] The three steps are not necessarily performed by the same person within the firm (e.g. the manager). For example, step (1) may be made within either the R&D or the marketing department, step (2) is a managerial task per se, and step (3) is an entrepreneurial task most closely associated to the owners of the firm.

> The most important features of genuine uncertainty are the inherent unlistability of all possible outcomes resulting from a course of action, and the complete endogeneity of the uncertainty. The first feature … is the basis of novelty or true surprise. This is in sharp contrast to the mere arrangement (or weighting) of known possibilities characteristic of neoclassical uncertainty. The second feature … is the origin of an ongoing market process that itself produces changes to which the system must adapt.

This quote captures well the essential problem of uncertainty and why it is important for economic science. The question for present purposes is whether Bayesian theory relies on the assumption that we know and can list all the possible outcomes. In other words, while the Bayesian theory might be better than the simple rational expectations model, it might still not fit *all* the intricacies highlighted by the Austrian theory of entrepreneurship. Discussing Bayesian probability theory, O'Driscoll & Rizzo note their main objection (1985: p. 4):

> The movement toward a "subjectivist" theory of probability in some areas of economics has no doubt been an improvement from our perspective. Yet most of this literature neglects a fundamental aspect of ignorance: the (perceived) unlistability of all possible outcomes. It is not merely that we do not know which possibility out of a given set will occur, but the set itself is unbounded. Subjective probability thus reflects subjectivism in its static form; while unbounded possibility sets reflect the essentially dynamic aspect of subjectivism. Real time and ignorance belong together.

This problem of unlistability is far less serious than O'Driscoll & Rizzo make it out to be. First on all, the real problem is not unlistability per se, but the fact that different people conceive of different (listable) sets of possibilities. In other words, the task of a theory of entrepreneurship is not to describe the mind of God from whose perspective the entire unlistable set of possibilities is contemplated, but to describe human beings who, for better or worse, only take small sets of possibilities into account at any given moment in time. To put it differently, one can deal with the important issue of epistemic heterogeneity without framing the entire matter from a "view from nowhere" perspective, in which epistemic heterogeneity no longer exists because all possible perspectives of human entrepreneurs (real or possible) have been merged into a single, all-encompassing, and "unbounded" perspective. The issue is thus simply to allow the list of possibilities to change over time, but at every moment it will be a "bounded" list.

Secondly, Bayesian theory actually deals explicitly with the problem of unknown unknowns (see for instance Bretthorst 1988: section 5.1; 1990; Jensen and Nielsen 2007: chapter 6). Bayesian theory deals with this issue by simply adding to the set of possibilities an additional "unknown" item. As Bretthorst put it (1988: pp. 55-6): "To say that we confine ourselves to the set [*X*] is not to assert dogmatically that there are no other possibilities; we may assign prior probabilities … which do not add up to one. … Then we are assigning a prior probability … to some unknown proposition, SE = Something Else not yet thought of." By taking this path we can show why the problem of unlistability is not that serious. It can be easily proved that, by adding SE to the set of possibilities, the *relative* probabilities of the *known* alternatives *remain* *unchanged*. This means that the rational decision made by taking the unknown unknowns into consideration will always be the same as the one made without taking them into consideration. In other words, as long as one keeps an open mind that one may be missing something, the issue of objective unlistability has no behavioral effects.

Finally, contrary to O'Driscoll & Rizzo claim, the Bayesian literature is actually quite concerned with what can aptly be called "subjectivism in its dynamic form". *It couldn't have been otherwise*: Once one understands that the numeric values of all probability distributions depend on one's assumptions, i.e. that, in general, $p(x|A) \neq p(x|B)$, one is naturally led to the problem of assessing the truth of those assumptions. The Bayesian "search theory", which, in the Bayesian sense of the word, is concerned with the creation, change and evaluation of the hypothesis space, is anything but static. Here is for instance how Jaynes (1985) summarizes it:

> (1) Think hard about the appropriate hypothesis space. Look for some symmetry/invariance property. (2) Try out your best choice. If the desired kind of useful results appear, then well and good—there is no evidence pointing to a different hypothesis space and you are done—at least for the time being. (3) If you get unsatisfactory results, then, if you are convinced that all relevant constraints have been taken into account, this is evidence that Nature is using a different hypothesis space than yours. Go to step (1).

Similarly, Gull (1988) notes that

> [t]he real art it to choose an appropriate "space of possibilities", and to date we have no systematic way of generating it. … [I]n many problems one has no guarantee that our choice is right in any final sense, and this feeling of ambiguity has led to much soul-searching. I feel (along with Jaynes 1986) that our aims should be different. We should not seek a "final truth" in our hypothesis space, but use our common sense to capture enough structure of the real problem being solved so that we can make useful predictions. If the predictions are useful, then that is an indication that the hypothesis space is good enough for now, without prejudice to the possibility of revising it later. If the predictions are not good, this is not a disaster, for we then have learnt that the hypotheses have to be reformulated and the ways in which our predictions are wrong may help us to do this.

I stipulate that this is exactly how entrepreneurs, as well as managers of established businesses, think as well. If one applies this search "algorithm" not to the process of discovering and correcting scientific models of nature (as Jaynes and Gull do in the quoted papers), but to the process of discovering new profit opportunities, one ends up precisely with Kirzner's concept of entrepreneurship and with the Kirznerian view of the market process.

In the economic case, the hypothesis space refers to assumptions one makes about consumer demand and about the various possibilities of combining existing resources in novel ways (of creating new "recipes" [Romer 1993]). Entrepreneurship thus involves a process of updating one's representations about the world (including the social world), an update which leads one to make *different estimations of risk* as compared to one's competitors who rely on different representations. Entrepreneurship does not thus involve merely a search within an objectively given, unique structure of probabilities, known and accepted by all, and within which actors move mechanically driven by maximizing expected profit. While it is true that we can model all actors as maximizing expected profit,[3] each of them acts based upon a slightly different representation of the world, and thus on different subjective probability distributions. Entrepreneurship in the Kirznerian sense does not refer to one's drive toward maximum expected profit (that's merely "management" as Sarasvathy (2001) would put it), but to the fact that one tries to improve one's representation of the world and thus be "alert" to opportunities that others don't notice (i.e. opportunities that other's judge as highly implausible due to their somewhat flawed hypothesis space). I thus think that it's safe to say that the Bayesian picture I'm presenting here includes entrepreneurial discovery.

6 Why there is no automatic tendency toward common knowledge

Is Bayes' formula mechanical? O'Driscoll and Rizzo (1985: p. 77) claim that "techniques such as modification of probabilities based on Bayes' Theorem permit only deterministic changes. Given the occurrence of a certain subsequent typical event, there is only one way the probability weights can be altered." This claim rests on a common misunderstanding stemming from not paying enough attention to how the probability distributions are actually generated. Bayes' formula seems mechanical only if one takes for granted that the new information (often qualitative) has already been transformed into a numeric probability estimation. However, it is precisely in this process of turning information into numbers that is the tricky part. Moreover, this process doesn't depend solely on the "data", but it also depends on one's *other* assumptions (see Annex for a simple illustration). Two persons with different background assumptions will usually code the same piece of information into different numeric estimations of the likelihood factor (with which the prior probability is then updated).

There are two interpretations of Bayes' formula. On one hand, it is understood as a *mechanical model of learning*. From this perspective, people are passive agents bombarded by information and Bayes' formula models the way in which they incorporate this information in their decision processes (which are also mechanically driven by expected utility maximization). On the other hand, it is understood as a *cognitive tool* useful in the creative process of trying to understand the world. From this perspective, people actively try to improve their hypothesis spaces and they use Bayes formula as an essential component in this model selection process, allowing them to explore the consequences of different possible choices of hypotheses.

It is this second interpretation which is the correct one, because the first one mistakenly assumes that information *can* be mechanically coded into probability estimations, while, in fact, such coding actually depends on one's assumptions about the hypothesis space. As Gull (1988) put it (emphasis added): "Bayes' rule is used to manipulate probabilities in the light of experimental data; [maximum entropy] is used

[3] although even this is can be doubted in the light of the heuristics literature (Gigerenzer et al. 1999)

to assign probability distributions given testable information. However, *it is up to us to choose a hypothesis space that is suitable for our problem*, and this not only requires us to assign an appropriate measure in the space of possibilities, but to define a range of allowed values for any parameters involved." Note that this is precisely the same as O'Driscoll & Rizzo's statement about the objective unlistability of possibilities. Objectively speaking, the set of all possibilities is indeed unlistable, but that's not the real issue: the real issue for an entrepreneur is to find a usable set of possibilities (be it discrete or continuous), i.e. to find a particular framing of her problem that is going to provide enough accuracy and deliver reliable enough predictions *for her practical purpose*.

According to this latter interpretation, entrepreneurs don't *just* look out for different pieces of information, but they also *organize this information differently*, based on their different background assumptions. This has the important consequence that there is no necessary common knowledge, i.e. even if people are presented with the same set of basic empiric information $\{I, J, K, \ldots\}$ they don't necessarily draw the same conclusions from it: if they are relying on different background assumptions $\{X, X', X'' \ldots\}$, they don't compute the same posterior probability distribution (see also Samuelson 2004, for a more in-depth discussion of the problem of common knowledge). This is the simple explanation for why honest disagreements between rational people are possible (Cowen and Hanson 2007). Cowen & Hanson's dilemma exists only if one assumes that everybody codes the given information into probability distributions in the same way, but, as Gull has pointed out, there actually isn't a single objective way of doing this.[4] In our language, we can say that this is the place where the entrepreneurial element enters the picture.

The issue of honest disagreements among perfect Bayesians highlights that there is nothing in the theory of probability per se that determines the convergence in terms of the most fundamental background assumptions. The purely Bayesian convergence can exist only in terms of those estimations of probability that rely on already existing common background assumptions. In his review article, Samuelson (2004: p. 377) draws a similar conclusion from a simpler, set-theoretical perspective, pointing out that "two agents whose information and experiences are identical *in every conceivable respect* should have identical beliefs", but only as long as the so-called Harsanyi doctrine is adopted, i.e. that "we should work with models in which agents have common prior beliefs". Consequently, if we adopt the Harsanyi doctrine, "[a]ny differences in beliefs can then be traced to the effects of differing information or experiences acting on these initial identical beliefs". By contrast, the Bayesian model selection perspective focuses on the process of creating and changing the *assumptions*, rather than merely taking them for granted as part of a "doctrine". To put it as starkly as possible, the Harsanyi doctrine simply leaves Kirznerian entrepreneurship out of economic theory by fiat. As Samuelson has put it, it creates the "impossibility to agree to disagree".

[4] The rigorous form of Bayes' formula is the following: $p(x|I,X) = p(x|X)p(I|x,X)/p(I|X)$. However, it is often written in abbreviated form as $p(x|I,X) = p(x)p(I|x)/p(I)$, which masks to some extent the subjectivity involved, in particular it creates the illusion that the likelihood, $p(I|x, X)$, doesn't depend on one's background assumptions X, but describes instead an objective relation between the new data I and the variable of interest x.

But what is the consequence of the fact that the Bayesian mechanism doesn't mechanically lead to a convergence of assumptions? The only conclusion is that such convergence is the outcome of a *social process*. Indeed, the investigations into this process of creating new hypotheses (Vosniadou and Ortony 1989; Clarke and Primo 2012) support the view that the convergence to common beliefs, to the extent that it exists, should be seen as a social process. Along the same lines, Samuelson (2004: p. 278) also notes that

> [t]he importance of common knowledge in the "agreeing to disagree" result raises the question of … how any event might become common knowledge. … If we impose sufficient structure on the interaction between agents (including, for example, implicit assumptions that they understand and can draw appropriate inferences from what they hear), then we can model a process in which knowledge effectively becomes common knowledge.

This type of investigation leads outside the scope of the present paper, but let me point out that a simple model of this social process can be developed by noting that such a convergence has certain benefits and Boettke and Coyne's (2009) social entrepreneurs emerge to mediate it. Coyne & Boettke model social entrepreneurship in terms of Shelling points, which from the Bayesian point of view are determined by the prior probability. The Shelling points perspective makes it clear why it is often useful for people to have converging background assumptions: lack of common knowledge leads to various forms of coordination problems, and, as a result, an interesting type of "entrepreneurship in non-market settings" emerges, concerning one's ability to mediate such convergence. As Coyne & Boettke argue, politicians can be seen as such mediators.

7 Conclusion

The main task of this paper has been to show the deep implications of the Bayesian claim that all probabilities are conditional probabilities. It was shown that the Bayesian perspective allows us to draw the distinction between uncertainty and risk in a way that is not vulnerable to the legitimate types of criticisms to which Knight's original distinction has been subjected. Moreover, this alternative way of defining this distinction naturally leads to Kirznerian entrepreneurship. In fact, the idea of entrepreneurship can now not only be explained in a perhaps clearer way, but we can also show how to actually measure surprise. To answer the question in the title, probability theory can deal with entrepreneurship but only as long as we are using the Bayesian theory in which all probabilities are conditional. The Austrian critique of the use of probability theory in the social sciences, and economics in particular, is correct only if we see it as targeting the logical positivist version in which probabilities are assumed to be objective properties of the world.

The following issues have been addressed: the impossibility of having an objective list of all alternatives over which to compute an objective probability; the heterogeneity of representations among different people; the fact that subjectivity takes a dynamic form as a result of model selection; the creative and non-deterministic aspect of the process by which representations are changed (the art of choosing the space of possibilities, as Gull called it); Kirznerian alertness and Kirznerian error resulting from the fact that different people, using different representations, have different probability estimations of various opportunities for profit and risks; the fact that the emergence of common knowledge is always the outcome of a social process mediated by institutional and cultural rules and social-entrepreneurs, rather than an intrinsic consequence of rationality.

Acknowledgments I have received very useful feedback on earlier versions of the paper from Dragos Paul Aligica, Simon Bilo, Peter Boettke, Bryan Caplan, Robert Cavender, Chris Coyne, Thomas Duncan, Jesse Gastelle, Laura Grube, Dave Hebert, Ryan Langrill, Peter Leeson, Jayme Lemke, William Luther, Lotta Moberg, Kyle O'Donnell, Shruti Rajagopalan, Alexander Salter, Solomon Stein, Virgill Storr, Jessi Troyan, and an anonymous reviewer.

Annex: probability assignments and background assumptions

This annex provides a simple example illustrating (1) that different background assumptions lead to different probability assignments of the same event, and (2) the rational expectations prediction is inferior to the general Bayesian method.

Suppose we have a die with an unknown number of sides, and we know from a series of throws that "1" has appeared once, "2" five times, and "3" four times (this is our data). From a rational expectations perspective, the probability distribution is $p \times (x) = \frac{n_x}{n}$, where n_x is the number of times x happened, and n is the total number of draws (1+5+4=10). By contrast, from a Bayesian perspective, which incorporates all kinds of information, we know, first of all, that the die, as a three-dimensional object, has to have at least four sides. Furthermore, when we start with the principle of indifference assumption of equal prior probabilities, and we then update using Bayes formula and the given data, we arrive at Laplace's rule of succession: $p(x) = \frac{n_x+1}{n+N}$, where N is the assumed total number of sides (Jaynes 2003: eq. 18.44). The comparative results are presented in Table 1 and Fig. 1.

As it should be obvious, the Bayesian answers are superior to the rational expectations answer because the rational expectations answer *assumes something which is not actually given*, namely that it is *impossible* for side 4 to happen anytime in the future. The rational expectations perspective effectively bans us from taking into consideration the relevant information that we're dealing with a three-dimensional object.

Table 1 Probability Estimates from the Same Empirical Data Under Different Theoretical Assumptions

Side number	Probability distribution					
	1	2	3	4	5	6
Rational expectations	0.10	0.50	0.40	0.00	0.00	0.00
Bayesian, N=4	0.14	0.43	0.29	0.07	0.00	0.00
Bayesian, N=5	0.13	0.40	0.27	0.07	0.07	0.00
Bayesian, N=6	0.13	0.31	0.25	0.06	0.06	0.06
⋮	⋮	⋮	⋮	⋮	⋮	⋮

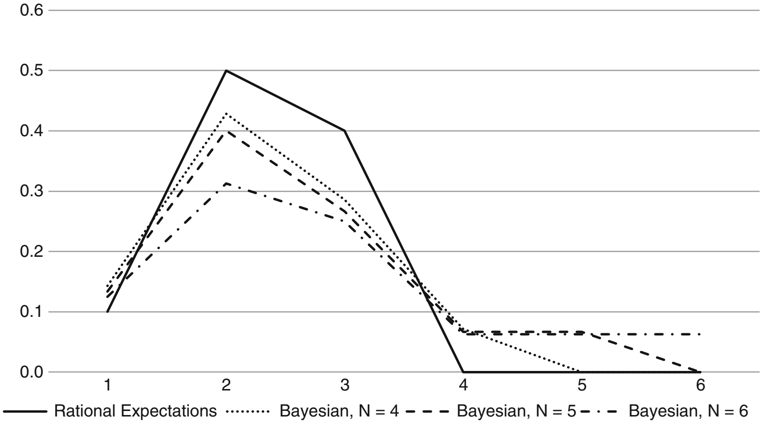

Fig. 1 Probability Estimates from the Same Empirical Data Under Different Theoretical Assumptions.

Moreover, the different Bayesian predictions, relying on different assumptions about *N*, highlight that the exact numeric values of the probability distribution, even in this case of a repeated experiment under homogenous conditions, critically depend on our theoretical assumption about the set of possibilities. Given that such complications appear even in the simplest examples, we see why Bayesians emphasize that there's no escape from the conditional nature of all probabilities.

References

Boettke, P. J. (1998). Rational choice and human agency in economics and sociology: Exploring the Weber-Austrian connection. In H. Giersch (Ed.), *Merits and limits of markets* (pp. 53–81). Berlin: Springer.

Boettke, P. J., & Coyne, C. J. (2009). An entrepreneurial theory of social cultural change. In V. P. Díaz (Ed.), *Markets and civil society: The European experience in comparative perspective* (pp. 77–103). New York: Berghahn Books.

Bretthorst, G. L. (1988). *Bayesian spectrum analysis and parameter estimation.* Berlin: Springer-Verlag.

Bretthorst, G. L. (1990). An introduction to parameter estimation using Bayesian probability theory. In P. F. Fougère (Ed.), *Maximum-entropy and Bayesian methods* (pp. 53–79). Dordrecht, Netherlands: Kluwer Academic Publishers.

Bretthorst, G. L. (1996). An introduction to model selection using probability theory as logic. In G. Heidbreder (Ed.), *Maximum entropy and Bayesian methods* (pp. 1–42). Dordrecht, Netherlands: Kluwer Academic Publishers.

Caplan, B. (1999). The Austrian search for realistic foundations. *Southern Economic Journal, 65*(4), 823–838.

Caplan, B. (2001). Probability, common sense, and realism: a reply to Huelsmann and Block. *Quarterly Journal of Austrian Economics, 4*(2), 69–86.

Clarke, K. A., & Primo, D. M. (2012). *A model discipline: Political science and the logic of representations.* Oxford: Oxford University Press.

Cowen, T., & Hanson, R. (2007). *Are disagreements honest?* Working Paper, George Mason University, Mercatus Center.

Crovelli, M. R. (2009). On the possibility of assigning probabilities to singular cases, or: probability is subjective too! *Libertarian Papers, 1*(26).

Gigerenzer, G., Todd, P. M., The ABC Research Group. (1999). *Simple heuristics that make us smart.* Oxford: Oxford University Press.

Gull, S. F. (1988). Bayesian inductive inference and maximum entropy. In G. J. Erickson & C. R. Smith (Eds.), *Maximum-entropy and Bayesian methods in science and engineering, vol. 1: Foundations* (pp. 53–74). Dordrecht: Kluwer.

Hayek, F. (1952). *The counter-revolution of science: Studies on the abuse of reason.* Glencoe, Illinois: Free Press.

Headd, B. (2003). Redefining business success: distinguishing between closure and failure. *Small Business Economics, 21*(1), 51–61.

Hirshleifer, J. (1970). *Investment, interest, and capital.* New Jersey: Prentice-Hall.

Hubbard, D. W. (2010). *How to measure anything: Finding the value of intangibles in business* (2nd ed.). New Jersey: John Wiley & Sons.

Itti, L., & Baldi, P. F. (2005). A principled approach to detecting surprising events in video. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1*, 631–637.

Itti, L., & Baldi, P. F. (2006). Bayesian surprise attracts human attention. In *Advances in neural information processing systems, vol. 19* (pp. 547–554). Cambridge, MA: MIT Press.

Jaynes, E. (1985). Entropy and search-theory. In C. R. Smith & W. T. Grandy Jr. (Eds.), *Maximum-entropy and Bayesian methods in inverse problems* (pp. 443–454). Dordrecht, Netherlands: D. Reidel Publishing Company.

Jaynes, E. (1986). Bayesian methods: General background. In J. H. Justice (Ed.), *Maximum-entropy and Bayesian methods in applied statistics* (pp. 1–25). Cambridge: Cambridge University Press.

Jaynes, E. (1988). How does the brain do plausible reasoning. In G. J. Erickson & C. R. Smith (Eds.), *Maximum-entropy and Bayesian methods in science and engineering, vol. 1. Foundations* (pp. 1–24). Dordrecht, Netherlands: Kluwer Academic Publishing.

Jaynes, E. (2003). *Probability theory: The logic of science*. Cambridge: Cambridge University Press.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian networks and decision graphs* (2nd ed.). Berlin: Springer.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91*(435), 1343–1370.

Keynes, J. M. (1921). *A treatise on probability.* London: Macmillan And Co.

Kirzner, I. (1973). *Competition and entrepreneurship.* Chicago: University of Chicago Press.

Kirzner, I. (1976). Equilibrium versus market process. In E. G. Dolan (Ed.), *The foundations of modern Austrian economics.* Kansas: Sheed and Ward.

Kirzner, I. (1980). *Perception, opportunity and profit: Studies in the theory of entrepreneurship.* Chicago: University of Chicago Press.

Kirzner, I. M. (1997). Entrepreneurial discovery and the competitive market process: an Austrian approach. *Journal of Economic Literature, 35*(1), 60–85.

Klein, D. B. (2012). *Knowledge and coordination.* Oxford: Oxford University Press.

Knaup, A. E. (2005). Survival and longevity in the business employment dynamics data. *Monthly Labor Review, 5*, 50–56.

Knight, F. H. (1921). *Risk, uncertainty, and profit.* Cambridge: Houghton Mifflin Company, The Riverside Press.

Lachmann, L. M. (1976). On the central concept of Austrian economics: Market process. In E. G. Dolan (Ed.), *The foundations of modern Austrian economics.* Kansas: Sheed and Ward.

Langlois, R. N. (1982a). *Subjective probability and subjective economics.* R.R. #82-09, New York University, C.V. Starr Center for Applied Economics.

Langlois, R. N. (1982b). *Entrepreneurship and knowledge.* R.R. #82-13, New York University, C.V. Starr Center for Applied Economics.

Laplace, P. S. (1840 [1951]). *A philosophical essay on probabilities* (6th ed.). (F. W. Truscott, & F. L. Emory, Trans.) Dover.

Lucas, R. E. (1977). Understanding business cycles. *Carnegie-Rochester Conference Series on Public Policy, 5*, 7–29.

Milgrom, P., & Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory, 26*(1), 17–27.

Mises, L. (1996 [1949]). *Human action: A treatise on economics* (4th ed.). San Francisco: Fox & Wilkes.

Mises, L. (1957 [2007]). *Theory and history.* Auburn, Alabama: Ludwig von Mises Institute.

O'Driscoll Jr., G. P., & Rizzo, M. (1985). *The economics of time and ignorance.* Oxford: Blackwell.

Phillips, L. D. (1970). The 'true probability' problem. *Acta Psychologica, 34*, 254–264.

Romer, P. (1993). Economic growth. In D. R. Henderson (Ed.), *The concise encyclopedia of economics.* Indianapolis, IN: Liberty Fund, Inc.

Samuelson, L. (2004). Modeling knowledge in economic analysis. *Journal of Economic Literature, 42*(2), 367–403.

Sarasvathy, S. D. (2001). Causation and effectuation: toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review, 26*(2), 243–288.

Sarasvathy, S. D. (2003). Entrepreneurship as a science of the artificial. *Journal of Economic Psychology, 24*(2), 203–220.

Sarasvathy, S. D., & Dew, N. (2007). *Without judgment: An empirically-based entrepreneurial theory of the firm.* Working Paper 84, George Mason University, Mercatus Center.

Schutjens, V., & Stam, E. (2006). *Starting anew: Entrepreneurial intentions and realizations subsequent to business closure.* Research Paper ERS-2006-015-ORG, Erasmus Research Institute of Management.

Shane, S. A. (2008). *The illusions of entrepreneurship: The costly myths that entrepreneurs, investors, and policy makers live by.* New Haven & London: Yale University Press.

Suppes, P. (1986). *Probabilistic metaphysics.* Oxford: Blackwell.

Vosniadou, S., & Ortony, A. (Eds.). (1989). *Similarity and analogical reasoning.* Cambridge: Cambridge University Press.

Wu, B., & Knott, A. M. (2006). Entrepreneurial risk and market entry. *Management Science, 52*(9), 1315–1330.