Elie Ayache

From Within

What is the smile problem? How can it be interpreted? Do people really understand it?

In the following, I will say what the smile problem is, and I am not sure everyone will follow me. This is how deeply and how long I've been thinking about the smile problem. I've pursued thinking about it in directions exactly opposite to the quantitative one, as you will see. You wouldn't be asking, at this stage, 30 years after the October market crash, what the smile problem is, if the answer were ordinary. The tools that it took me so long to develop, in order to understand the smile problem, are not quantitative tools. They are philosophical tools, which question and dig into the categories of thought: writing, trading, time, formalism, etcetera.

The smile problem isn't something that happens to the Black–Scholes–Merton (BSM) model from outside. It is not a *falsification* of the BSM model. The smile problem isn't that BSM assumes the underlying price process to be lognormal and that it incidentally happens in reality, externally to BSM, that the process is different, i.e., admitting of stochastic volatility and jumps. The smile problem is produced from inside.

If it were limited to the outside descriptive view, the smile problem would have been solved a long time ago. In that view, it would simply appear that the underlying probability distribution admits of a third and a fourth moment; derivatives would be evaluated in some martingale measure that is equivalent to the real one. As a result, implied volatility smiles would manifest themselves relative to BSM. That's a statistician's, econometrician's, smile problem; and I don't think it is the one that interests us.

In another view of the smile problem, which may seem difficult, at first, to distinguish from the previous one, BSM says that derivatives are redundant and do not trade in their own independent market, when in fact they do. This seems equivalent to relaxing lognormality; however, in this view, before even unfolding the argument that lognormality implies perfect dynamic replication and therefore redundancy of the derivative, the notions of *trading* the underlying asset and of *trading* the derivative intervene first. In the previous view, no trading was mentioned, as it was only a matter of *valuation* of lotteries, given a random generator and nonarbitrage. In the previous view, states of the world are assumed, which only *coincide* with prices of the underlying asset, as well as a probability distribution overlying them, and the valuation equally of the underlying asset and of the derivative is considered, with the only difference that one lottery – the underlying asset – admits of a trivial payoff, and the other lottery – the derivative – of a more complex one.

The reason why the second view is in fact different from the first one is that we are, from the start, considering the derivative as tradable, if only to conclude that it will be redundant as a matter of fact. But if the derivative is tradable, then we could as well argue that its own trading process should be given from the start, in parallel to that of the underlying asset. How could BSM presuppose that the derivative value is only a function of time and of the underlying price, in order, then, to find that the derivative is perfectly dynamically replicable by the underlying asset, and therefore redundant? Why is the derivative 'value' not made additionally a function of another trading variable, specific to the volatility market, or trivially a function of the derivative *price*? As a matter of fact, I for one believe that as soon as the derivatives

market is considered, the whole register becomes incompatible with the notion of time series of the underlying price and the corresponding econometrics

But how is the derivative even considered in the first place? If all that we have at our disposal, at the beginning, is the trading pit of the underlying asset, and the corresponding stochastic representation, how could we ever access the notion of the derivative, let alone its market?

* * * *

If, by the BSM world, we understand a world that is constituted by an asset whose price (ideally) follows Brownian motion because it is trading freely, then how is a derivative written on that asset ever introduced? If all that exists and that is of concern, in that world, is the trading of the basic asset, how do we form the thought of a derivative asset in that world? Think of the trading activity of the basic asset; think of the traders immersed in its trading pit; think that the only measure of time that they have is the next transaction they will execute, and the only measure of space is the next change of price; think that the only dynamics they know is that of the auction process, that the only reason they have to buy the asset is that they anticipate that its price will go up, and the only reason they have to sell it short is that they anticipate that its price will go down, keeping in mind that the upward and downward movements themselves are caused by nothing else than the buying and selling pressures (Keynes's beauty contest); then how, in such a limited and, as a matter of fact, perfectly closed world, can we engage the dimensions of time and space that are implicit in the act of writing the derivative? How can we understand the logic that would be encoded in the written derivative – namely, that somebody would *later* have to pay a certain amount of money, only if a certain threshold is reached? How can we understand the delay and the condition that the derivative involves, both in time and in space, when all that happens is the present and instant trading of the underlying asset?

The beauty (and the mystery) of a trading asset is that the movement of its price is random by necessity and that the price it is currently trading at is by necessity the one at which to purchase the lottery ticket yielding its future possible prices as outcomes. Risk-neutral pricing was invented for the sole purpose of identifying the asset's current trading price with the expected value (or present value, to adopt the actuarial language) of the asset's future payoff (which is none other than its future price) under a given probability measure. This intrusion of probability and the backward schema it introduces are here just paving the way for the nonarbitrage principle. Suddenly, we need to get ahead of ourselves and go look in the future in order to identify the possible states of the world. Suddenly, the trading price is no longer the trading price and is no longer the sole driver. The road becomes now completely mapped in front of it and the price becomes a value. Trading the asset freely in the market becomes equal to evaluating, by backwardation, the lottery that its future prices will constitute.

It is in such a framework, in which the time of the market has been perverted and its trading force has been suspended, that derivatives can be considered as written beforehand. When we step outside the trading pit and look at the underlying asset price externally, as the ordinary random generator that it has now become, the thought of prewritten derivatives becomes possible. For they are now to be considered as mere lotteries, in equal measure with the underlying asset itself. Now, the nonarbitrage principle can be properly framed and the order of thought can take place, in which future outcomes are conceived prior to present value. Crucial to its enunciation, as we said, is the notion of states of the world. It may seem as if the underlying asset is still being exchanged in its pit, but the trading activity (what we have called the *trading force*) is no longer fundamental; it becomes mere plaster that is covering an existing structure. A different time than the volcanic time of the trading pit has suddenly been conceived. Somebody has suddenly articulated the thought that the underlying asset's prices now unfold as a time series by virtue of which a statistical problem can be posed. Before we even consider the time series of underlying prices and envisage its statistical analysis – determining, for instance, whether the statistical distribution is stationary or not, and envisaging ways that its moments may be estimated – we avail ourselves of the corresponding notion of time, in which the series is staged.

It may look, from the point of view of physics, that there is but one category of time, and that it is in the 'same time' that the trading action is taking place and the time series of prices is eventually registering its numbers. However, I maintain that the two time perspectives are different and even

It may look, from the point of view of physics, that there is but one category of time, and that it is in the 'same time' that the trading action is taking place and the time series of prices is eventually registering its numbers. However, I maintain that the two time perspectives are different and even incompatible with each other

incompatible with each other. This is really our problem. This is really the smile problem. The market is not a physical entity. In the realm of physics, there is but one reality and one time.

Setting aside the details of relativity theory which makes time dependent on the frame of reference, the nature or the register of time that physics considers is one and not many. The trajectory of an elementary particle can be correlated with the flight of a bird, with the speed of a train, or with the growth of a cell. They all take place in the same reality; therefore, this common reality can act as the common cause explaining their correlation, even though their 'correlation' may, after analysis, never be reduced to fewer explanatory variables than the trajectories themselves. This is another way of saying that they share the same reality and, if not in reality, at least in fiction a writer can propose to detect common patterns ruling their evolution.

By contrast, the market does not happen in physics. By saying this, I am not pointing toward psychological time and suggesting that, because the market is, above all, a human affair, the psychology of a trader immersed in a pit is different from the psychology of a statistician who considers, from outside the pit, the full extension of the time series of prices. Neither am I embracing a sociology of finance, in which the society of derivatives traders, armed with statistical knowledge and a technology – BSM – superior to those of the basic trader of a basic underlying asset, can step in and help to create the new reality of the derivatives markets, and shape the corresponding world. My register is semantics. I want to understand what our best quantitative theories of the market and their formalisms are saying or meaning to say. Perspective is crucial. Chances are that the market, when it is looked upon from above, totally empirically, is just chaos in which derivatives trade as well as underlying assets.

* * * *

The smile problem is so fundamental in the financial markets that I would like to argue that the inversion that characterizes it occurs only in quantitative finance (or in derivative pricing), and in no other quantitative science which might equally be using a formalism and its interpretation, formulas and their results. We start with a formula or a formalism for trading (of the underlying) and, instead of ending up with a theoretical valuation (of the derivative), we end up with trading again (of the derivative). How could trading ever be a result? Trading should always be a given and never be result. We are only ever, and forever, immersed in trading, so how did we ever step outside the trading pit in order to deduce something from it (a theoretical result), and, more amazingly, once we had gone outside, how did we ever manage to get back inside and obtain trading again? As trading is immanent and there is no way of looking at trading from outside, we can say that the smile problem is the characterization of the trading; it is trading eternally returning; it is the *formula* of trading.

So, the smile problem is the deviation that re-establishes what has always been the given – namely, the existence of derivatives market prices, as opposed to values

The smile problem is not what people usually think it is – a deviation from the hypothesis of lognormality that underlies BSM. No matter how complex we make a certain model in order to account for the smile problem, this model will admit of a smile problem in turn. By that, we mean that the derivatives of the next level that it is intended to evaluate theoretically, in this case exotic options (for vanilla options are now trivially evaluated by calibration of the smile model against their market prices), will in turn deviate from their prescribed values and exhibit 'smiles.' So, the smile problem is the deviation that re-establishes what has always been the given – namely, the existence of derivatives market *prices*, as opposed to values.

But how is that even possible? How is the first deviation from theory happening? How is the derivative trading happening in BSM? What is its mechanism?

Because we are not describing a statistical or empirical reality that takes place in the register of empirical time, because perspective is important and we are keen on following the exact wording of the formalism in order to delimit exactly its interpretation, the BSM formula and what lies outside the formula are both needed. Strictly speaking, BSM doesn't mention the derivative. All it does is consider the trading process of the underlying asset and instruct us what premium to invest in it at the start, and subsequently how to dynamically trade it in a self-financing way, in order to manufacture predefined contingent payoffs at certain predefined maturity dates. The writing of the contingent claim (or derivative), and consequently its trading as independent asset, can only happen outside the BSM formula. For this reason, we need to maintain the formula in order that there remains an outside in which the derivatives market would occur. This is the reason of the smile problem, which amounts to keeping the BSM formula, while twisting it.

The smile problem is characteristic of the market. It is not about to go away. It is like the problem of measurement in quantum mechanics. It is inherent in the logic of the science itself. It is connected to the deepest categories in the science, to its archaeology: what writing means, what trading means.

* * * *

We shouldn't look at the market from above or even step outside the market. We should start from the trading pit of the underlying asset and see the smile problem *properly and strictly* emerge from there. The smile problem emerges because we start from a confined and complete trading pit, that of

the underlying asset, and end up with something unexpected, properly with an event: the trading of the derivative. BSM is the complete view of the trading pit of the underlying asset, and more so than Brownian motion, because it combines the randomness of the trading price with the total latitude of the trading size (the dynamic strategy). It is from inside this complete and perfectly closed world that the smile problem has to emerge, as we have said.

Trading the underlying asset – or, generally, trading anything – is contrary to any conception of the value of that thing, or concept of its valuation. We buy and sell the thing only in anticipation of a rise or a decline of its price; and its price rises or declines because it is bought or sold. From this groundless activity, *the volatility of the price emerges as the only value*. If asked what the meaning or the concept of a free exchange is, if asked what the 'value' of the market is, we should answer: "*volatility*."

There is a difference between the time register of empirical reality, in which the time series of the asset's prices is inscribed, and the time register of semantic certainty. Nobody has *witnessed* the underlying asset price following Brownian motion in empirical time. Rather, the concept of the market (or unpredictability of the price evolution at any time scale) is recognized and volatility is posited: "Let volatility be σ." In mathematical notation, this translates into Brownian motion. In this reading, Brownian motion is not a model of a time series of prices (as if the market was just another random generator). It is not a model of the reality of the market, but a model of its *meaning*. The time series of prices, which follows from Brownian motion, is a later consequence. It is a consequence of the formalism. A further consequence of the formalism and a further semantic translation of the volatility of price of the underlying asset is the premium starting with which we replicate a contingent payoff by a self-financing dynamic trading strategy involving the underlying asset and the money account. In other words, BSM.

The premium to replicate the contingent payoff is no less a *value* than volatility. It belongs on the same semantic level. It is of a different nature altogether than the price of the underlying asset. It has a different meaning altogether. There is a conceptual certainty in the constitution of the premium which is equal to the conceptual certainty of volatility (what we have called the *value* of the market). Options market-makers using BSM are potentially armed with a certainty of volatility and not, as everyone believes, with its uncertainty or stochasticity. This speaks of the extraordinary transmutation that has to take place in order that the premium to replicate a contingent payoff (this value, this certainty) becomes *the traded price of a contingent claim*. What will have to vary is the concept: from contingent payoff to contingent claim, from the completion of the formalism of the trading of the underlying asset to the exit from it, an exit which cannot but (which is programmed to) fall back into the trading pit, under the form of the traded contingent claim.

The smile problem comes neither from the external view, in which a random generator triggers lotteries which everybody confuses with derivatives, nor from the view from above, in which everything trades indistinctly. The proper view of the smile problem is that of writing, and consequently of trading, the contingent claim from inside BSM.

The smile problem is specific to BSM, inasmuch as BSM is *the completion and closure of the trading of the underlying asset*. The smile problem is the same as the concept of implied volatility. Implied volatility is only accessible to traders engaged in dynamic replication, because they alone replicate the contingent payoff, and consequently they alone write the contingent claim from whose price they alone can now infer implied volatility. It is they who invent the writing, and hence the trading, of the contingent claim.

Once the contingent claim becomes traded alongside the underlying asset, the dynamic replication strategy generalizes to include the two of them, and BSM generalizes to a stochastic volatility model and eventually to a jump-diffusion model with stochastic volatility, as the variance swap eventually starts trading independently of the replicating strip of options. As we said, the smile problem is transferred from one level to the next in an endless chain, and it is in the generalized sense of the term that we should understand that it is specific to BSM.

Implied volatility has to be separated from *informational efficiency*, or from the idea that the volatility that is implied from the market prices of derivatives eventually aligns itself with the real volatility of the underlying price, as if in a learning process

* * * *

There needs to be something tight and binding for there to be a smile problem. There wouldn't be a smile problem in the general description by martingales. There is a smile problem because we ascend from the trading pit of the underlying asset to conceptualize volatility and, *in the same movement and same conceptual certainty*, replicate the contingent payoff. There is a smile problem because this then coincides with writing the contingent claim – because this *invents* its writing – and because we are immersed back in the market. There is a smile problem because of the conjunction of both the certainty of replicating the contingent payoff (which is a semantic certainty) and the subsequent slippage to the contingent claim.

Of course, the smile problem arises because of the unexpected trading of the contingent claim. However, what makes it tradable is the replication of its payoff, or the *invention of its writing*, in the environment of trading the underlying, *which we have never left*. It is certainly not the independent process of stochastic volatility – or, in other words, econometrics – that makes it tradable. We have never left the trading pit of the underlying and the corresponding trading force. We have never turned into statisticians and the contingent claim has never been considered as written in advance.

There is no smile problem without implied volatility, which should be understood first. Implied volatility has to be separated from *informational efficiency*, or from the idea that the volatility that is implied from the market prices of derivatives eventually aligns itself with the real volatility of the underlying price, as if in a learning process. Informational efficiency supposes that the derivative has long been trading, and stably enough for the information to propagate. (It is a tribute to empirical reality, from which it hasn't liberated itself.) Whereas implied volatility is produced on the spot, as a total surprise: suddenly the derivative is written; suddenly it trades. It is because we are surprised by its price that the price enters as an input. It is the surprise that creates the reaction of inverting the formula, not the empirical learning process. Only because it has been produced in the void, as a total surprise relative to the formalism, does the price of the derivative invert the formula. The surprise takes place in the conceptual realm and implied volatility has nothing to do with the empirical realm.

We completely miss the dynamics of the market if we do not attend to the notion of recalibration. Recalibration, or *the mutation of the contingent payoff into a contingent claim*, is the cause of the smile problem. The contingent claim and its price have *literally* jumped outside BSM. Hence, they do not call for a reformulation of BSM that would solve the problem. They call for a generalization of BSM that would carry the problem to the next level. Hence the persistence of the smile at any level and the persistent usage of BSM (or its correct generalization). The smile problem is not a stage between two models, happening within the same descriptive reality, for instance, BSM and a stochastic volatility model. It is a stage between two levels of reading, or registers.

There is no smile problem in empirical reality and in the empirical trading of the underlying asset and the derivative. It poses itself to the formalism. We need to retain the formalism (i.e., a certain formula) in front of the market of derivatives. Hence, the smile, or the tension in the formula. Retaining the formalism, or the formula, is not only for the purpose of nonarbitrage. We need to express formally the derivative character of the derivative, if only to express its delta. The formalism of martingale evaluation grants nonarbitrage, but forsakes the trading force.

The smile problem is the fundamental problem of quantitative finance because it issues from time

The trading of the derivative cannot be formalized. In the formalism of martingales and nonarbitrage evaluation, there is a derivative but no trading. In the formalism of BSM, there is trading, but no derivative. Better to keep the trading, and to create the derivative from inside.

* * * *

It should be part of the smile problem that it won't go away; that every stage of trading should engender its own smile problem. The smile problem doesn't come from quantitative Brownian motion (i.e., volatility is not constant); it doesn't come from the time series dimension. Rather, it is essential to the smile problem, and it should be part of understanding what the smile problem really is, that it should come from the *qualitative* Brownian motion, from Brownian motion as the meaning of the market and not from Brownian motion as its quantitative description, from something pure, inside which quantity and measurement haven't broken. It is essential that the smile problem should come from the materialization into a tradable contingent claim of the contingent payoff that has just been replicated, and not from anywhere else.

The smile problem is the fundamental problem of quantitative finance because it issues from time. Time is the fundamental dimension in the market. There is the time of the time series of prices, the time of the stochastic evolution of prices, or the time of statistics. And there is the time of the end of time, the time in which the concept of the market is formed and the semantic certainty of volatility is reached, the time of the mathematical postulation in which the dynamic replication of the contingent payoff is conducted. These different registers of time depend on perspective: how do I step outside of the series and start looking at it and expecting it statistically; how do I engage in dynamic replication and entertain the semantic certainty of volatility in order to set up the trading strategy?

Science in the market has, first of all, a problem of perspective. What is the model really looking at; or, what amounts to the same, what is its time register? This is because tools in the market are used to make it, not just to look at it from outside. Sociology only got interested in the mechanism by which the formula shaped its market or its world. But sociology doesn't say what the model or the formula was initially intended for, what its world was originally, what its ontology is. According to it, there was a formula, naively describing a world, and it is accidentally that the society of its users prepared and shaped the world for it.

There wouldn't be a smile problem without the leap in the void, which is outside the formalism of the trading of the underlying and of the replication of the contingent payoff. The void that is located between the contingent payoff and the contingent claim is the reason why the twist, which is characteristic of the smile, won't go away. The twist of the smile problem is a twist over the void. The formula of replication of the payoff is twisted beyond its limit, onto the level where it cannot

reach. The smile reports implied volatilities, and it is important to see why implied volatilities cannot and should not align themselves with the (real) volatility parameters of the underlying process.

There wouldn't be a smile problem in the martingale valuation framework. In the martingale valuation framework, the underlying market is just a random generator. States of the world have been established and there is no longer trading or the market. We have exited the market and its pit, and all we are reporting are numbers that are generated randomly, which only used to coincide with prices. They have the same numerical values as prices, but the register is now different. All that is now in effect is the theory of nonarbitrage valuation. We select an equivalent martingale measure and we value the derivatives as a result. Conversely, if the values of derivatives, construed as mere lotteries, were given, we could reverse-engineer the martingale or risk-neutral measure. This is a legitimate inverse problem. The same formula is being used in one direction, then in the opposite direction.

In the framework of replication of contingent payoffs, there wouldn't be a smile problem either. If the premiums to replicate contingent payoffs were given, the dynamics of the underlying assets that were used to replicate them could be reverse-engineered too (numerically, it is the same inverse problem as the one before). This remains a mathematical formula that is being exploited in one direction, then in the opposite direction.

Notice that in both cases, the fact of being given the values of derivatives (the lotteries) or the premiums to replicate contingent payoffs is a thought experiment. If somebody were to give us those values or those premiums, then we could formally invert the formula. To know the values or the premiums is formally the same as knowing the underlying dynamics; a formula lies between the two. Just as we can formally imagine the dynamics, we can imagine that somebody is giving us the values or the premiums.

In reality, however, it is the prices of contingent claims that are given, not values of lotteries or premiums to replicate contingent payoffs. They are given by the market. There is no longer a formula leading from the underlying dynamics to the prices of the contingent claims, because the void has cut inside the formula. What we are feeding back in the formula, in order to invert it, is no longer homogeneous with the formula; it is no longer of the same nature as the formula. The formula is used to output values of derivatives (lotteries) in the martingale framework, or premiums to replicate contingent payoffs in the replication framework; and now we are feeding back prices into it. This is the smile problem. The smile problem is not an inverse problem.

Even before the option price deviates from the premium to replicate its payoff, the fact of feeding back a market price in the formula, rather than a premium, is already posing the smile problem. If volatility were to vary, the premium to manufacture the contingent payoff would vary; accordingly, a different premium would mean a different volatility; this is the meaning of the two being related by a formula; this is the reason why both the direct and the inverse problem are, in this case, meaningful. However, in the real case, what varies is not the value of the volatility corresponding to a variation of the premium. What varies is not quantitative. What varies is that the contingent payoff becomes a contingent claim. The whole market introduces itself in between, in the void.

If somebody tells us that the market price of the contingent claim is such, why are we not confident that the volatility of the underlying price must be such?

If somebody tells us that the market price of the contingent claim is such, why are we not confident that the volatility of the underlying price must be such? Why are we not really confident about inferring volatility – a statistical parameter – in this way? People who think there is still one formula relating the underlying process and the price of the contingent claim, in order to invert the former with the latter as input, must believe that the dynamic replication strategy involving the underlying asset replicates contingent claims *in the market*. However, if you think about it, no algorithm that replicates a contingent payoff can force the market price of the corresponding contingent claim to align itself with the premium required for replication. This is because of the incompatibility of registers of time. To consider the *price* of the contingent claim is to consider the market – the same market in which trading of the underlying is performed for the dynamic replication strategy and not a later market. Although the formalism can't tell the difference, considering the contingent claim as tradable is incompatible with the register of statistics.

To write and subsequently to trade the contingent claim is to move back inside the trading pit after a semantic ascent that has already broken away from the time dimension of the time series. The reality of the market is the reality of a *price series*: the vanilla option price is invented inside a pit where only the underlying price existed at first, then the exotic option price is invented inside a pit where only the vanilla price existed, and so on and so forth. The smile problem is only the consequence of realizing that this price series – or the process of recalibration – is not a time series.