"*Constructive Skepticism*" Volume 3 – Notebook #I: Model Risk

Chapter 6: "*Does Your Model Look Like a Mini-Fig*"

Let's retrace our path since Chapter 1 before we start this chapter. These reading notes about Model Risk started with the observation that the Myth of "*Iron and Spinach*" had iconic qualities for the reading of research papers. So much so that these reading notes use the word "*Spinach*" to describe "*Things we think unquestionably true but look ambiguously false after asking a few questions*".

Chapter 2 provided added the background details on the Myth of "*Iron and Spinach*" that gave us a two-part framework to look at models in research papers from the perspective of "*Constructive Skepticism*": (i) The Measurement Problem, and (ii) The Preference Problem.

Chapter 3 continued to develop the framework with the "*Statistical Meaning*" of the Measurement Problem, and the "*Practical Meaning*" of the Preference Problem. This revealed the critical importance of "*Dimension Reductions*" in modeling, thus identifying the model risk of all model risks: "*Confusing a Part for the Whole*".

Chapter 4 described "*Statistical Illusions*" associated with the Measurement Problem, and started to show how we can work around them by structuring quantitative models as combinations of Random Variables, Probability Distributions, Change [Growth] Formulas, Stochastic Processes, Realized Trajectories [Evidence Data or Simulated], Calculated Averages & Growth Rates, and Transformation Functions.

Chapter 5 described the "*Roughness*", Complexity & "*Randomness*" associated with the Preference Problem. This highlighted the importance of "*Fractals*" in working with the Preference Problem, and contrasted with the traditional "*Smooth, Simple & Continuous*" functions commonly used to solve the Measurement Problem.

In the spirit of CTRI's charter to translate the findings of research papers in plain English and clear mathematics, this chapter shows that modeling requires making "*Scaling Choices*" with effects that we can understand inuitively by looking at LEGO® Mini-Figurines ("*Mini-figs*"). This metaphor leads us to adding a third model risk, "*Does Your Model Look Like a Mini-fig*", to a list that already includes "*Spinach*", and "*Confusing a Part for the Whole*".

<u>The Modeling “*Biases*” of the “*Mini-Fig*”</u>

The iconic LEGO® brick, a 3-D rectanglar prism with a 2x4 array of studs on the top face, speaks to the problem of “*Scaling Choices*” in plain English: A rectangle crushes things down in two of its three dimensions. This “*Scaling Choice*” becomes a problem that we can sense intuitively because LEGO® bricks – as do the equations in quantitative models – do the work of mathematical functions that transform original evidence data into final modeled results.

The brick functions as a Affine Transformation of the original into the model, and our sense of aesthetics, associated with the Preference Problem - based on our ability to recognize the “*Fractal*” nature of reality - tells us intuitively when these brick-based models do not look right. Solutions to such aesthetic rejections include changing the basic measurement unit used to build the model. For instance, we can move down from the iconic 2x4 brick to the Unit Stud Brick: A single stud Brick of the same height as the iconic brick.

Human life moves on a scale where inches & feet make intuitive sense. As we move down to fractional dimensions these human-based measurements lose their real-life relevance, and become difficult to use because they scale up and down in “*Powers of Two*”, and we find it easier to make such calculations in “*Powers of Ten*”. This led to the creation of the Metric System in general, and in this case its preferred use in defining the measurements of the Unit Stud Brick. Based on papers such as ***Tom Alphin***’s June 2016 “*Matters of Scale*” in Bricks Magazine #13, the Unit Stud Brick has:

- Sides of 8.0 mm for the square in 2-D
- Height of the resulting rectangular prism in 3-D: 9.6mm
- Height of the stud: 1.6 mm
- Diameter of the stud: 4.8 mm
- Distance between studs: 3mm
- Thickness of brick wall, called the “Pip”: 1.5mm

These measurements represent the “*Axioms, Assumptions & Hypotheses*” for building models with the Unit Stud Brick. However, models built from such Unit Stud Brick continue to show the structural, dimension-crushing “*Bias*” of the Iconic 2x4 Brick, and the consequences of this “*Bias*” extend to the shape of LEGO®’s Mini-fig.

Using the Unit Stud Brick as the basic model unit defines the dimensions and shape of Mini-figs as follows:

- Height: 4 Unit Stud Bricks placed on top of one-another = 9.6 mm x 4 = 38.4 mm plus a single Stud of 1.6 mm on top of the head, resulting in a total of 40 mm, resulting in a mid-point scaling ratio to an average human being of 1:42
- Width: 2 Unit Stud Bricks placed side-by-side = 8 mm x 2 = 16 mm, resulting in an empirical scaling ratio of 1:25 according to ***Alphin***

The “*Scaling Choices*” of the Unit Stud Brick turn the “*Mini-fig*” into a mix of apples- and oranges between the dimension for height & the dimension for width. The requirement to have two Unit Stud Bricks side-by-side makes the “*Mini-fig*” proportionately wider than it should. Thus, the end-result departs noticeably from an intuitive sense of expected proportions for a human being.

Earlier in these reading notes, we used the word “*Spinach*” as a metaphor to describe “*Things we think unquestionably true but look ambiguously false after asking a few questions*”. Now, we introduce the use of the word “*Mini-fig*” as a metaphor do describe “*The intuitive perception of aesthetic dissonance that comes from intrinsic dimensional scaling bias in a model*”.

In order to address dimensional scaling “*Bias*”, LEGO® modelers can: (i) Keep the rectangular prism dimensional scaling constraints , and move down the scale to find the smallest possible unit in order to mitigate the visibility of the problem with finer grain resolution, (ii) Keep the rectangular prism dimensional constraints model, and find combinations of basic elements that can build-up into a Cubic Assembly, or (iii) Abandon the rectangular prism dimensional scaling constraints, and create a new Unit Stud Cubic Brick.

Moving down the scale takes us first to the Unit Stud Plate, and then to the Pip. Moving up the scale takes us to ***Xander Henderson***’s Cubic Assembly developed to increase the accuracy of his models of fractals structures such as the Menger Sponge.

First, the Unit Stud Plate has one-third the height of the Unit Stud Brick (9.6mm/3 = 3.2mm = 2 Pips), and mitigates the “*Perception*” the Preference Problem by its ability to create fractal-looking landscapes, as described by ***Antonio Bellon*** in his series of 5 articles about his “*Modular Integrated Landscaping System*” (MILS). This series first published in 2012 in a publication called Hispa Brick Magazine® ran from issue 013 to issue 017. ***Bellon*** focus on the Unit Stud Plate set many best-practices standards for creating “*Fractal*” looking landscaping with LEGO® bricks, including the following:

- Model trains can climb a grade at the rate of one Unit Stud Plate per 16 studs of distance,
- Modeled hills grow most realistically at the rate of one Unit Stud Plate per stud of distance, and
- Modeled mountains grow most realistically at the rate of one Unit Stud Brick per stud of distance.

However, the MILS standards for “*Roughness*” in the modeling of such landscpaes do not solve the “*Mini-fig*” dimensional scaling bias, nor do they solve its impact on the modeling of smoother surfaces, or the scaled reduction of volumes. We see what we understand. ***Bellon*** made us understand how to build and see the “*Fractal*” aesthetic primitives in landscapes built with LEGO® bricks. However, the Unit Stud Plate does not provide a sufficient level of granularity or “*Resolution*” to create matching aesthetic primitives for realistic models of manufactured objects based on smooth and continuous functions.

Thus comes the second option, the Pip, a mesurement derived from the thickness of a brick's wall. The Pip appears to be the smallest unit possible for modeling with LEGO® bricks. The Pip enables the use of creative building techniques - such as "*Stud Not On Top*" (SNOT) – to better approximate the smooth and continuous curves of manufactured objects. Additionally, Pips simplified model planning because it turned the external decimal metric measurements of the Unit Stud Brick into integer measurements intrinsic to the model as follows:

- Sides of 8.0 mm = 5 Pips
- Height of 9.6mm = 6 Pips
- Height of the stud: 1.6 mm
- Diameter of the stud: 4.8 mm = 3 Pips
- Distance between studs: 3mm = 2 Pips

However, use of the Pip accelerated the development of an already growing catalog of specialized parts. While the exact number of LEGO® parts does not seem to have been officially released by the company, its numbering system evolved over time from 4 digits (Thousands of potential parts) to 7 digits (Millions of potential parts) in order to track the different types of elements in various colors & part designs over their production history. This level of growing complexity shows the intrinsic costs as well as a structural source of "*Randomess*" associated with the irreducibility of a "*Bias*" problem in a modeling unit.

One can marvel at how a Measurement Problem difference of 1 Pip (1.5 mm) between the two surface dimensions, and the height dimension compounds such a seemingly small flaw in the basic "*Unit of Account*" into visible Preference Problems at various levels of scale. The iteration of small flaws into mis-shaped signals makes the connection to Chaos Theory.

This leads to questions about finding the right balance between the choice of basic unit, model detail and model minimalism:

- While ***Bellon's*** MILS found the right level of detailing for modeling the "*Roughness*" of fractal landscapes, what is the right level of detailing for modeling manufactured objects built from smooth and continuous functions?
- At what point does adding details exceeds the "*Resolution*" of the model?
- What intrinsic structures limit the real-life insights that one can take from a model?

These questions bring us to the third option that requires the development of a new basic unit. For instance, modelers need a symmetrical brick in all there dimensions to create models of "*Fractal*" mathematical objects. These mathematical objects would otherwise not come out right if modelers used the asymetrical "*Biases*" of the previously encountered basic "*Units of Account*".

Based on an April 16, 2014 blog post by ***Xander Henderson***, titled “*Lego® Fractals*”, creating the smallest possible Cubic Unit - given the foundational “*Bias*” of the basic Unit Stud Brick & Unit Stud Plate – requires the following combinations:

- Four Unit Stud Bricks with sides of 8mm arranged as a 2x2 square create a combined Unit Cube with sides of 16 mm
- The four Unit Stud Bricks have a common height of 9.6mm to which Henderson adds two Unit Stud Plates of 3.2 mm each for a total height of 16mm that makes is a cube

This combined Cubic Unit brick eliminates the foundational “*Bias*”, and achieves proportional reduction in all three dimensions but at the cost of a loss in granularity. We can see the size of this loss this by comparing the size of this Unit Cube to the size of the “*Mini-fig*” as follows:

- Unit Cube = 16 mm x 16 mm
- “*Minig-fig*” = 16 mm x 40 mm

Henderson’s Unit Cube takes 40% of the front surface area of the “*Mini-fig*”. Solving the “*Bias*” problem created another problem that precludes realistic modeling with such Unit Cubes. This Unit Cube works as a specialized basic unit to solve specific mathematical objects as contrasted with scaling down human-scale physical objects.

As a short aside to keep the length of this chapter short, combining modeling techniques can lead to “*Units of Account*” that trigger unexpected behaviors. Starting in 2005 with ***Steve Hassenplug***, modelers combined the LEGO® soccer ball, the best-practices standardization of MILS, and the use of all existing construction techniques ranging from the basic block to the Pip to create a new modeling domain called the Great Ball Contraption (GBC). GBCs look like self-contained manufacturing lines that combine all manners of moving, lifting, throwing, and sorting 14mm balls in a consistent, coherent way with repeatable high performance.

In this case, the soccer ball with a 14 mm diameter becomes the symmetrical - but spherical instead of cubic - “*Unit of Account*” that removes the “*Mini-fig*” from the picture. GBC model “*Processes*” instead of physical objects. Thus, the spherical “*Unit of Account*” creates stationary increments that merge the dynamic model with its “*Task Environment*”.

The model and the “*Task Environment*” become one display. GBC models create its own context. One can think of them as train models that do not requires a landscape to put them in context. GBC models have 4 dimensions: The three physical dimensions plus a time dimension made visible by the dynamics of “Motions”. GBCs make me think of Ergodicity Economics.

However, the GBC building techniques keep the foundational asymmetries of Unit Stud Brick & Plate. Thus it may be that the never-ending appeal of building with LEGO®

comes from the psychology of the "*Orienting Response*". Its structural "*Bias*" captures our attention, and prevents us from ever getting it right. This built-in tension between the Measurement Problem and the Preference Problem motivates us to try, and try again.

This trip into LEGO® nerdiness gave us examples, metaphors, and questions that we can use to review models in Financial Economics, Behavioral Finance, and Retirement Planning, including:

- "*Does Your Model Look Like Mini-fig*"?
    - What dimension reductions did a Modeler make to simplify the Measurement Problem of reality into a specific model?
    - What scaling choices, made by the modeler across these dimensions, impact the Preference Problem for the user of the model?

- What "*Axioms, Assumptions & Hypotheses*" lead to Measurement and Preference Problems cannot be solved within the model?

- What levels of complexity and "*Randomness*" can we expect to find in models built on matrices with thousands of dimensions that reflect physically unconstrained "*Mind Exuberance*" such as Large Language Models (LLMs)?
- How can models of greater complexity than LEGO® bricks convince us that their view of reality does not look like "*Mini-figs*"?

- At what point should we move from a "*Mini-fig*"-looking model to another model based on better "*Axioms, Assumptions & Hypotheses*"?

The rest of this chapter looks at the back-and-forth between the ***Heuristics & Bias Program*** and the "*Fast & Frugal*" ***Heuristics Program*** from this perspective.

<u>What Models from the Heuristics & Bias Program Look like "*Mini-Figs*"?</u>

As shown earlier with the discussion of modeling with LEGO® bricks, a model fits in a landscape, a "*Task Environment*". The model and the landscape may or many not share the same dimensions – for instance a model may have a time dimension as it moves across a static landscape. Futher, a model may not have proportionally consistent dimensions. Finally, models and landscapes may or may not share consistent dimension reductions, and scaling choices, and this brings us to the differences between the ***Heuristics & Bias Program*** and the "*Fast & Frugal*" ***Heuristics Program***.

In his many books ***Gigerenzer*** gives frequent and explicit recognition to ***Herbert Simon*** in general, and ***Simon's*** concept of the "*Scissors*" in particular [*The dual interaction of an individual's mental abilities and the characterisitics of their "Task Environment", as the individual confronts multiple goals and conflicting choices*] as the inspiration for the development of "*Ecological Rationality*".

Gigerenzer expands the debate beyond ***Simon***'s "*Scissors*" to apply his concept of "*Ecological Rationality*" to both "*Large World*" problems, and "*Small Worlds*" problems. For instance, and as described in Chapter 4, pages 52 to 55 of ***Gigerenzer's*** 2023 book "*The Intelligence of Intuition*" illustrate the proper matching, and the proper use of "*Small Worlds*" statistics in the context of their relevant "*Task Environment*". It also shows that our intuitions about "*Randomness*" under conditions of "*Uncertainty*" in the "*Large World*" [instead of "*Risk*" in the "*Small Worlds*"] achieve this proper matching and use automatically:

- "*Rational*" decision-makers make good decisions when they use the "*Tools*" of the ***Logic & Statistics Program*** to match the circumstances of ***risky*** "*Task Environments*", and
- Intuitive decision-makers make good decisions when they use the "*Processes, Heuristics & Algorithms*" of the "*Fast & Frugal*" ***Heuristics Program*** to match the circumstances of ***uncertain*** "*Task Environments*".

In real-life, "*Small Worlds*" risky "*Task Environments*" become the exception, and "*Large World*" uncertain "*Task Environments*" become the rule. However, heuristics develop over time, and on the basis of individual experience. The proper use of heuristics by comes from "*Long Experience*".

Gigerenzer's catalog of validated heuristics and their matching "*Task Environments*" makes it possible for the rest of us to develop, and to trust our intuitions faster than we would otherwise could. It also makes it possible to implement these heuristics in the form of computer algorithms for the greater benefit of AI development.

However, two historical schools of thought give different meanings to the word heuristic:

Starting in 1905 with ***Albert Einstein***, continuing with ***Max Wertheimer, George Polya***, and formalized in the 1950s by ***Herbert Simon*** & ***Allen Newell***, the first school of thought defined the word heuristics as follows:

- "*... incomplete but highly useful*" views,
- "*... methods such as "looking around" to guide search for information.",*
- Methods "*to find a proof*" [as contrasted with analysis] " *for checking a proof.",* and
- Formal models to "*limit large search* spaces".

During the 1970s, researchers from this first school of thought, that spans several disciplines, formalized algorithmic models, and demonstrated the adaptive value of:

- *"lexicographic rules, elimination-by-aspect, and equal-weight rules*",
- *"rules of thumb (their term for heuristics)*" used by the animals studied by behavioral biologists, and
- "*Heuristics*" used by artificial intelligence researchers for "*problems that logic and probability*" cannot solve such as computationally intractable problems.

On the other hand, and also in the 1970s, Psychologists took a different direction, and formed the second school of thought called the ***Heuristics & Bias Program*** that:

- "*… became interested in demonstrating human reasoning errors*",
- "*… used the term heuristic to explain why people make errors*.",
  Replaced formal *"models of heuristics by general labels such as "availability" and later, "affect".*", thus "*Unlike in biology and AI, heuristics* [in Psychology] *became tied to biases, whereas the content-free laws of logic and probabilities became identified with the principles of sound thinking*."

The ***Heuristics & Bias Program*** started to diverge from ***Daniel Benoulli's*** "*Expected Utility Theory*" (EUT), and ***Jimmie Savage's*** "*Subjective Expected Utility Theory*" when ***Ward Edwards*** started "*Behavioral Decision Theory*" in 1954. ***Edwards***' students pursued a program of empirical demonstrations of observed divergences from the "*Predictions*" of EUT. These empirical violations from theory included:

- Preference reversals, and
- Framing effects.

Edwards' findings led ***Amos Tversky*** & ***Daniel Kahneman*** to start their formalization of the ***Heuristics & Bias Program*** in 1971. This work includes "*Prospect Theory*" (PT) as a proposed alternative to EUT. PT started in 1979 with ***Kahneman & Tversky's*** paper "*Prospect Theory: An Analysis of Decision under Risk*". This paper documented "*Effects*" violating EUT's "*Independence of Irrelevant Alternatives*" (IIA) assumption [*Binary preferences do not change with the introduction of a third option*]. "*Effects*" that violate IIA include:

- "*Risk Aversion*" [*A preference for smaller variance in outcomes*],
- "*Loss Aversion*" [*Losses are felt more intensely than gains*], and
- The "*Endowment Effect*" [*The framing of decisions as gains or losses affects choices*].

PT changed the shape of ***Bernoulli's*** utility curve from a concave function over positive values to an S-curve function over positive and negative values as an "*As-if*" model – instead of a causal model - to explain these "*Effects*" within the decision-making framework of the "*Rational*" ***Logic & Statistics Program***.

Nicholas C. Barberis' 2013 paper "Thirty Years of Prospect Theory in Economics: A Review and Assessment" points out that PT is widely known as: *"… the best available description of how people evaluate risk in experimental settings.*" He also finds it curious that: *"… there are relatively few well-known and broadly accepted applications of prospect theory in economics*." Finally, he explains the divergence by contrasting the central idea of PT: *"… people derive utility from "gains" and "losses" measured relative to a reference point.",* with the observation that *"… in any given context, it is often* *unclear how to define precisely what a gain or a loss is, not least because* ***Kahneman & Tversky*** *offered relatively little guidance on how the reference point is determined."*

In the discussion part of the paper, **Barberis** points out that proposed PT applications remain "… *hypotheses in need of more testing*" before we even ask the next question: "*If people avoid annuities, "overpay" for initial public offerings, or go to casinos because they evaluate risk with prospect theory, does that mean that these behaviors are mistakes?*" This means that while PT may have started as a "*Repair Program*" for EUT, it eventually devolved into a "*Maintenance Program*" of the ***Logic & Statistics Program*** by limiting itself to psychological explanations of the "*Puzzles, Paradoxes & Anomalies*" of the normative, "*Rational*" recommendations. Further, these explanations blame the "*Decision-Makers*" with irrationality instead of examining alternatives to the "*Axioms, Assumptions & Hypotheses*" that support the "*Rational*" prescriptions.

Ad-hoc psychological narratives cannot turn "*As-if*" models into causal explanations. For instance, PT could not address the observed "*Four-fold Pattern of Preferences*" [*Given very high or very low probabilities, empirical decision-makers switch their approach to risk*], leading to an evolution from the "*Value Functions*" of 1979 Prospect Theory (PT) to the "*Weight Functions*" of 1992 Cumulative Prospect Theory (CPT).

In 1992 with ***Tversky & Kahneman's*** paper "*Advances in Prospect Theory: Cumulative Representation of Uncertainty*" introduced a fix to PT's problems with the introduction of Cumulative Prospect Theory (CPT). The paper provides a solution to the "*Four-fold Pattern*" using data derived from a computer-based laboratory experiment with 25 graduate students from Berkeley and Stanford. Note the small sample size, and compare it to the large flight of psychological fancy that it claims to support.

This pattern of low-powered studies recurs with uncomfortable frequency in ***Behavioral Finance***, leading to issues that range from the non-reproducibility of results from such low-powered studies, to cherry-picked data, and even outright fraud. This seemingly endemic problem even led to public retractions of famous results such as "*Priming*" due to the Nobel Prize winner and lead author having *"...placed too much faith in underpowered studies*."

Tversky & Kahneman summarize their results for CPT with an inverse-S curve of cumulative probabilities on the X-axis, and cumulative decision weights on the Y-axis to show that the decision weights of experimental decision-makers diverge from the authors' expected probabilities in situations with very low or very high probabilities, and solve the problem with the following psychological explanations:

- First, "*Decision-Makers*" with a near-certain gain, perhaps fearing disappointment, shift from risk-prone to risk-averse,
- Conversely, "*Decision-Makers*" with a near-certain loss, perhaps hoping to avoid a loss, shift from risk-averse to risk-prone,
- Second, "*Decision-Makers*" with a small chance of loss, perhaps fearing the pain of losing, shift from risk-prone to risk-averse, and
- Conversely, "*Decision-Makers*" with a small chance of gain, perhaps hoping to receive a gain, shift from risk average to risk-prone.

In the discussion part of the paper, ***Tversky & Kahneman*** point out that “*Despite its greater generality, the cumulative functional is unlikely to be accurate in detail.*”

Can Ergodicity Economics Fix Models that Look Like “*Mini-figs*”?

Coming at the problem from another direction, ***Ole Peters*** replicates CPT’s inverse-S curve of cumulative probabilities vs. cumulative decision weights by mapping the cumulative differences between two probability distribution functions. Thus, probability weighting reflects the difference between the location (mean), scale (variance), and shape (distribution) of two decision-making models. This logical, mathematical mechanism sheds light on ***Tversky & Kahneman*** caveat about CPT’s “*lack of accuracy in detai*l”, given the likely presence of observable decision-making models other than their use of the “*Rational Investor*” model.

In a 2020 paper titled “*What are we weighting for? A mechanistic model for probability weighting*”, **Ole Peters, Alexander Adamou, Mark Kirstein & Yonatan Berman** provide a mechanistic explanation as a potential replacement for the probability weighting solution of cumulative prospect theory. The solution hinges on the difference in estimates of expected values and variances between the model of a “*Disinterested Observer*” (DO) and the model of a “*Decision-Maker*” (DM).

According to ***Peters, et al***., the iconic inverse-S curve that compares cumulative DM decision weights by cumulative DO probabilities is the mechanical consequence of the following observations:

- The greater uncertainty of a DM modeling uncertainty over a specific, single life-cycle trajectory, as contrasted with
- A DO modeling uncertainty over the ensemble of possible trajectories.

As explained by ***Peters, et al***., “*The inverse- curve does not mean that “Probabilities are re-weighted”. It means only that experimenters and their subjects have different views about appropriate models of, and responses to, a situation.*” Thus, what looked like an insightful story of deep psychological flaws with individual “*Decision-Makers*” now looks like a banal story of run-of-the-mill measurement “*Bias*” on the part of the “*Disinterested Observers*”.

We see what we understand, and researchers used low-powered studies that we too-good-to-check in support of their preconceived ideas. The presence of low-powered studies in support of “As-if” models in support of normative prescriptions gives us yet another source of model risk, and this brings us back to the “*Mini-fig*”: What looked like an external feature of reality, now looks like the result of squashed perspectives springing from problematic internal foundations.

Michael Polayni’s “*We can know more than we can tell*” also means that we can tell when something does not look right. The next chapter will focus on Hypothesis Testing as a modeling technique that creates models that do not look right, and how this leads to the introduction of “*Invisible Gorillas*” as a fourth model risk on a list that includes “*Spinach*”, “*Confusing a Part for the Whole*”, and “*Your Model Looks like a Mini-fig*”.