Puzzled by idealizations and understanding their functions[1]

Uskali Mäki
University of Helsinki
Nankai University

Abstract
Idealization is ubiquitous in human cognition, and so is the inclination of being puzzled by it: what to make of ideal gas, infinitely large populations, *homo economicus*, perfectly just society, known to violate matters of fact? This is apparent in social science theorizing (from JH von Thünen, JS Mill and Max Weber to Milton Friedman and Thomas Schelling); recent philosophy of science analyzing scientific modelling; and the debate over ideal and non-ideal theory in political philosophy (since John Rawls). I will offer a set of concepts and principles to improve transparency about the precise contents of idealizations (in terms of negligibility, applicability, tractability, and early-step status) and their distinct functions (such as contributing to minimal modelling, benchmark modelling, and how-possibly modelling).

Keywords
Idealization, ideal type, ideal and non-ideal theory, minimal model, benchmark model, how-possibly model, negligibility, applicability.

1 Introduction

Mass point, frictionless plane, vacuum, ideal gas, infinitely large populations, immortal humans, perfect competition, perfectly informed agents, zero transaction costs. And so on. Across the history of natural and social science, idealizations of various kinds have prospered, and it appears science has made a great deal of progress thanks to – or despite of – them. Yet, there is persistent puzzlement about what to make of idealizations simply because they appear to be in stark conflict with the facts as we know them and because their useful functions are seldom fully transparent. Confusion and puzzlement have occasionally given rise to heated debates.

---

[1] The article derives from a keynote lecture delivered at the Asian Conference in the Philosophy of the Social Sciences, held at Nankai University, Tianjin, China, on 1-3 June 2019. The paper has also been presented at GATE, University of Jean Monnet, Saint-Etienne; and at TRIANGLE, University of Lyon 2. Thanks for helpful comments go to the participants of these events and to two anonymous referees.

In what follows, I will examine these puzzlements in the social sciences, through authors such as JH von Thünen, JS Mill, Max Weber, Milton Friedman, and Thomas Schelling; in contemporary political philosophy and the recent methodological debates therein around "ideal and non-ideal theory" since John Rawls and others; and in contemporary philosophy of science and the massive investigation into scientific modelling. There are interesting affinities and variations between these conversations as well as obvious opportunities for them to learn from one another.

My overall goal is to improve the transparency of idealizations in use. I will not argue for or against any particular idealization, but will outline a framework – a few organizing concepts and general principles – within which debates over any idealization can be fruitfully carried out. This requires clarifying the very idea of idealization; the alternative intended *contents* of various idealizations; and the distinct *functions* served by them.

2 What to make of extreme and obvious falsehood

I will provide some opening observations about the key components in the conundrum. I am suggesting that idealization is ubiquitous, and so is puzzlement about it. We need to be clear about what we are talking about.

Puzzlement.
On the one hand, there is puzzled *resistance* to idealization. One obvious background presupposition, explicit or implicit, is the principle that science should pursue the goal of discovering facts about the world. When faced with an idealization, one may conclude that this is not the way the world is, or at least this misses its richness. This judgement has sources and it has intensities.

In judging that an idealization does not get the facts right, one may draw from a variety of background information. In much of social science, commonsense observation plays major roles. Almost anyone, regardless of educational background, can tell that market

agents are not perfectly informed lightning calculators, just based on one's experience with human action and interaction, or on general cultural understanding of human life (cf. the idea of "phenomenological pressure" in Mäki 2013, 89-90). In contrast, it takes a little education in physics to know that nothing travels faster than light and that therefore the idealized concept of rigid body – that assumes that force is transmitted from one part of the body to another in no time – gets the facts wrong.

The puzzled judgement comes with various intensities. They may range from caution to suspicion to outright repudiation. More weakly, one may merely have stronger doubts about a theory or model that employs unadorned idealizations than about one that doesn't. More strongly, one may disallow idealizations as unacceptable. In this vein, Peter Blau stated, in the context of his account of bureaucracy, that "[s]ince generalizations about idealized states defy testing in systematic research, they have no place in science" (Blau 1956, 35). In the context of political philosophy, Onora O'Neill finds it objectionable that in idealization "much (too much) that is false of human agents is *added*" (O'Neill 1987, 56).

On the other hand, there are puzzled *justifications* of idealization. Faced with an idealization, and perhaps an objection to it, one may feel the need to provide a defense. The task is not easy and straightforward. Many of the intuitions invoked in these situations tend to make sense, but they are seldom very systematic and elaborate, and they may miss relevant possible defenses. Hence some attempted justifications may give rise to further puzzlements and confusions (as we will see).

The recommendable attitudes should derive from selective judgement that would be passed largely on a case by case basis. This will require understanding what each particular idealization says and does as well as what standards would apply in assessing its capacities for promoting the goals of inquiry. It is the purpose of this paper to facilitate such judgements.

Idealization.

Next consider what these hesitations, resistances and defenses are about. What is an *idealization*? There are two views in the literature that are not helpful. First, some authors characterize idealization simply in terms of falsehood: they are false elements in models and theories (e.g. Valentini 2009, 332, 338). Second, some consider idealizations simply in terms of deformation, failing to distinguish idealization from other deformational procedures and their outcomes, such as simplification, abstraction, omission, exaggeration, approximation, or generally "misrepresentation". Note also that the 'ideal' in 'idealization' does not just refer to the operation residing in the theorist's or modeler's *ideas.*

As I see it, idealization can be taken to have distinct features that dissociate it from the above two characterizations. First, if taken as factual propositions, idealizations are not just false, but deliberately so, associated by an *awareness of their falsity*. Their falsehood is obvious, and it is accepted for a purpose. Idealizations are strategic falsehoods, deliberately employed by modelers to achieve their goals. This means idealizations are *not errors*, mistakenly believed to be true. They are *not lies* either (contra Cartwright 1981), known to be false by their utterers but presented as true to the utterers' audiences to deceive them. And they are *not hypotheses*, submitted for critical scrutiny to determine whether they are true or false. In all such cases, a statement would be removed if revealed to be false, whether initially an error, lie, or hypothesis. By contrast, if an idealization is removed, it is not removed for this reason. Its falsity is known from the start, and if it will be removed, this will happen for more refined reasons.

Second, idealizations do not appear to be just false, but their ideality is also connected to some sort of *perfection*, often put in terms of *extreme* values of properties, such as zero and infinity. One idealizes by assuming friction or transaction costs to be zero, or populations to be infinitely large or their members to be infinitely lived. Sibling operations such as mere simplification and exaggeration would not qualify as

idealization without being associated with ideality as perfection.[2]

In some cases, ideality is conceived normatively. It designates normative ideals that are to be recommended and pursued. Their desirability may have moral, political, aesthetic, or other non-epistemic origins. Idealizations in political philosophy are often intended to articulate normative ideals.

Why idealize?

Many authors consider idealization a matter of self-evidence in first-rate science. Doing science in accordance to what Leszek Nowak calls the "idealizational conception of science" is taken by him to have been the best practice from Galileo and Marx and beyond (Nowak 1981).[3] Yet in scientific practice itself, puzzlement, confusion, and diversity of opinions prevail. The issues are far from settled, so we need to go on asking questions and trying to answer them. What to make of idealization? Why idealize? Let me provide a first quick and rough idea of the reasons and possible justifications for using idealizations in our epistemic pursuits. There are at least three sets of such reasons (cf. Potochnik 2017):

1. The world is immensely rich in attributes and complex in relationships.
2. Our cognitive capacities are limited.
3. Our cognitive interests are multifarious.

Reasons 1 and 2 alone imply the necessity to simplify and idealize in our cognitive ventures. My main focus in this essay will be on reason 3, itself sufficient for the importance of idealization or its equivalents for our epistemic aspirations.

---

[2] It is naturally also possible to adjust the terminological conventions. We could talk about idealizations in a narrow or strong sense, denoting assumptions that are put in terms of extreme perfection; and in a broad or weak sense, comprising (some) other deformational assumptions.

[3] While Nowak celebrates Marx as the pioneer of the idealizational method in the social sciences, I have analysed von Thünen as an earlier exponent and practitioner who additionally was exceptionally self-aware of the functions of the idealizations he used.

Functions of idealization

I will next provide some preliminary elaboration of reason 3. The key is to understand the *functions of idealization*. I propose dealing with this in two steps. On step one, we recognize the *experimental moment* in modelling, exhibiting its *root function* (Mäki 1992a, 2005a). Idealizations function like the controls in laboratory experiments, adopted and manipulated to control for "disturbances" or "impurities" in the situation at hand. The modeler assumes certain factors to be ideal and thereby *isolates* some other features of the situation for closer examination. Just like experimental controls (that keep a factor constant or absent) are not lies or errors, idealizations aren't falsehoods in those ways.

On step two, we ask what *further function* is served by the ("experimental") isolation that is effected by means of idealizations. There are several such further functions, but here I will focus on just three important ones, plus a fourth that blends the other three with a normative dimension. Further elaborations and illustrations will be provided as we proceed, but here are the preliminaries:

> [1] Idealizations may serve the function of *isolating a causal mechanism* or a simple dependency relation. Many models in natural and social sciences are *minimal models* in this sense: they help focus on the structure and operation of one mechanism at a time, in isolation of all other mechanisms and conditions that in actual circumstances may also shape the behavior of the phenomena under investigation.
>
> [2] Idealizations may help envisage *a benchmark* that is not directly used as a description of anything real. The actual world is considered a set of *deviations* from the benchmark. By seeking to explain those deviations as the *explananda* one aspires to organize one's understanding of the world.
>
> [3] Idealizations may be used for exploring *possible causal scenarios*, for explanation or for design. A model employing idealizing assumptions may isolate a possible causal mechanism, one that possibly has brought about, or is possibly causally sustaining (or possibly will cause or sustain) a phenomenon to be

explained (or a system to be designed).

[4] Idealizations may serve the function of exploring *normative scenarios* that involve ethical or political ideals. This typically involves some or all the other functions 1-3 as well.

Contents of idealizing assumptions

It is a mistake to treat idealizations as (false) *assertions* or statements about the world (contra O'Neill 1987 and Jones 2005). Consider the assumption of a closed economy, relying on idealizations such as a country's Exports = 0 and Imports = 0. This composite idealization can be used for making several other claims (see Musgrave 1981; Mäki 2000, 2012; Hindriks 2006):

> *Negligibility:* That the country's exports and imports > 0 is negligible given the purpose for which a closed-economy model is being used.
>
> *Applicability:* The model applies to economies that are closed or whose deviation from full closure is negligible.
>
> *Tractability:* Assuming a closed economy enhances the mathematical or theoretical tractability of the model.
>
> *Early-step:* The closed economy idealization is part of an early version of the model that will be relaxed in later versions.

None of these make the (false) claim that an economy is closed. These are alternative altered versions of a straight idealization that assumes some extreme perfection. These altered versions can be used for explicating the functions of idealizations and thereby for alleviating puzzlements around them. Note that a straightforward idealization, if interpreted literally and considered a truth-apt claim, is known to be false, while these altered versions have a chance of being true (Musgrave 1981; Mäki 2012).

Puzzlements again

We can now be a little more specific about the nature of puzzlements about idealization. There are at least three sources of puzzlement. First, one may not have understood the

identity and contents of an idealization correctly. It may be unclear what exactly is being talked about, and what exactly what is talked about says, and about what – e.g. whether it is supposed to function as a negligibility claim or an early-step promise. Second, and in part related to the first, one may not have understood the *essential functionality* of an idealization, i.e. that it is used to serve some further function and so should not be considered a factual assertion and appraised as such. The third kind of puzzlement arises when one has not understood the *specific apposite function* of an idealization, and instead judges it as serving some other function for which it is unfit. For example, one may mistakenly judge (what should be considered) a benchmark model as a minimal model and therefore draw incorrect conclusions in one's assessment.

The sections to follow aim to offer tools for improving the transparency of idealizations in terms of their contents and functions. Note that idealizations, supposing they are explicitly stated or at least available for being explicitly stated, are relatively transparent themselves – especially when contrasted with the massive amount of silent and implicit omission that is unavoidable in modelling. However, idealizations as explicitly stated are not transparent enough, since their formulations typically do not reveal their intended (or required) contents and functions. Their transparency needs upgrading.

3 Minimal model idealization

A major function of idealization is to help to isolate a minimal causal fragment embedded in a larger causal structure.[4] Some of the more puzzling and controversial cases belong to this category.

---

[4] There is a slight difference between my use of the phrase 'minimal model idealization' and that of Michael Weisberg (2009). Weisberg seeks to distinguish "three kinds of idealization" in modelling, of which "minimal model idealization" is one, while I don't take them to be different kinds of idealization – but rather different kinds of function that idealizations may perform. Till Grüne-Yanoff's (2009) use of 'minimal model' is much further away. It deviates from the mainline use adopted here. On his usage, minimal models are not representational, while on the standard reading, they do represent, even if something minimal.

J.S. Mill 1836

J.S. Mill's (1836/1844) famous thoughts about how political economy treats human behavior exemplifies this function. Only a small selection of behavioral dispositions is isolated by economic theory:

> "It does *not treat of the whole of man's nature* as modified by the social state, nor of the whole conduct of man in society. It is concerned with him *solely as a being who desires to possess wealth and who is capable of judging* of the comparative efficacy of means for obtaining that end. ... It makes *entire abstraction of every other human passion or motive*; except those which may be regarded as perpetually antagonizing principles to the desire of wealth, namely, *aversion to labour*, and *desire of the present enjoyment of costly indulgences*." (Mill 1844, 137-138; italics added)

Economic theory thus isolates the rational maximization of wealth (together with two counteracting dispositions, namely aversion to labor and desire of present enjoyment) from all other motives that are implicitly assumed to have zero strength. One can then draw inferences about what happens in such an imagined economic world. In response to suspicions about employing such a narrow view of human motivation, the obvious justification is to argue that the isolated dispositions are the most powerful and stable ones in the economic context. This was Mill's own (realist) line. We can put it in terms of *negligibility*: even though other motives are real and may have an impact, their impact is negligibly small given the purposes for which the selection of relevant motives is being put. This involves an *applicability* component: the idealizations are supposed to be adequate when applied to the study of the economic domain; and one may expect the presence of other motives to be more negligible in this domain than in some non-economic domains.[5] Such negligibility claims ("other motives are negligible") and

---

[5] "With respect to those parts of human conduct of which wealth is not even the principal object, to these Political Economy does not pretend that its conclusions are applicable." (Mill 1844, 139)

applicability claims ("the theory best applies to the economic domain") may be true even if agents do not quite rationally maximize and even if they are actually driven by numerous other motives as well, in whatever domain, economic or otherwise.

Milton Friedman 1953

It is not always appreciated that Milton Friedman's (1953) confused defence of the maximization assumption, more than a century later, is similar to that of Mill. Friedman defended the assumption against criticisms based on empirical research that had concluded that business firms do not actually seek to maximize their expected returns. He famously generalized this into an encompassing argument that the unrealisticness of assumptions in economic theory does not matter. This has been interpreted by many commentators to imply an instrumentalist, anti-realist conception of theory and model. I have argued this is a mistake: whatever Friedman's authentic philosophical outlook may have been, the essay (F53 for short) *can be* read as a realist statement (see e.g. Mäki 1990, 2009). This seems evident in the following passages.

> "A meaningful scientific hypothesis or theory typically asserts that certain forces are, and other forces are not, important in understanding a particular class of phenomena." (F53, 40)

This is in line with suggesting that unimportant forces are to be idealized away. The falsity of such idealizations could be accompanied by a true negligibility claim: those forces are there, but they make a negligible impact on the phenomena under study, thus idealization is justified. Here is another rendering:

> "It is frequently convenient to present such a hypothesis by stating that the phenomena it is desired to predict behave in the world of observation *as if* they occurred in a hypothetical and highly simplified world containing only the forces that the hypothesis asserts to be important." (Ibid., 40)

It is another mistake to believe that once an *as-if* formulation is used, one thereby

commits to an anti-realist position (echoed in popular phrases like "Friedman's instrumentalist as-if methodology"). Here we have a realist version of the *as-if*, used for expressing the idea that a minimal model ("a hypothetical and highly simplified world") isolates important causes ("containing only the forces that the hypothesis asserts to be important").

However, F53 is quite confused in its defense of unrealistic assumptions. One example is the conflation between nothing but the truth and the whole truth in envisioning "a completely 'realistic' theory of the wheat market" (ibid., 32) against critics of the maximization assumption (see Mäki 2009, 97-98). Another comes from the use of a presumed analogy with assuming a vacuum in Galileo's law of freely falling bodies. Analogies often illuminate, but this one obscures (see Mäki 2009, 99-101). F53 argues that if vacuum is fine, then so is profit maximization. However, vacuum (air pressure = 0) is an idealization with functions similar to those of assuming magnetic forces = 0, all other forces = 0, and the radius of the Earth = ∞ (flat Earth) in the case of Galileo's law; and assuming that the strength of all other motives except the maximization motive = 0 in the case of theory of competitive firms. The adequate analogy that is left unexamined by F53 is between profit maximization and gravity, both isolated by idealizing other factors away.

Galileo's law is useful for illustrating the roles of negligibility and applicability considerations. F53 does not employ these labels, but it provides arguments that manifest the concepts. F53 says Galileo's law is mostly applicable to falling cannon balls in atmospheric conditions, but not to falling feathers. The reason is that the deviation of the actual behavior of the cannon ball from what Galileo's law predicts is negligible for many purposes, whereas the deviation in the case of a feather is not. This is to say the falsity of the vacuum idealization about actual atmospheric conditions is negligible in some cases (to which it is then applicable) but not in others.

J.H. von Thünen 1826/42
JH von Thünen's famous and still influential model of the isolated state not only isolates

the state from the rest of the world, but it also isolates a small mechanism that is believed to make a major impact on the distribution of land use (see e.g. Mäki 2011). The model is dependent on many highly idealizing assumptions: the area is flat; the soil is uniform in fertility; the area is uniform in climate; there is just one a point-like Town in the middle of the area; the agents are perfectly informed rational maximizers; and so on.[6]

Given these assumptions it is possible to derive a neat land-use pattern of concentric rings, the famous Thünen rings. The function of the various idealizing assumptions is that of controlling for several causally relevant factors (such as mountains, rivers, fertility and climate differences, other towns, biases in decision-making, etc), excluding or neutralizing them, cleaning them away, as it were. Thünen thereby prepares a purified stage where just one purified mechanism shapes the outcome. Idealizing assumptions help isolate a major mechanism and its characteristic way of operation. This is the *Thünen Mechanism* that links distance from the Town with the emerging land use pattern, operating through land values and transportation costs. Closer to the Town, land values are higher and transportation costs lower, and further away from the Town, land values are lower and transportation costs higher. Given their assumed rationality, producers in the model make optimal location decisions, balancing these two distance-

---

[6] A fuller set of the model's assumptions would include the following.
1. The area is flat, there are no mountains and valleys (vertical dimension = 0).
2. There are no navigable rivers or canals.
3. The soil is throughout capable of cultivation.
4. The soil is homogenous in fertility.
5. The climate is uniform across the state.
6. There are no contacts between the area and the outside world ("Isolated State").
7. At the center of the plain there is a town, a geometric point.
9. There are no other towns in the area.
10. All industrial activity takes place in the town.
11. All market interactions between the producers in the town.
12. Transportation costs are directly proportional to distance and to the weight and perishability of the good (no roads, no preservation technology, delivery by oxcart).
13. All prices and transportation costs are fixed.
14. Production costs are constant over space.
15. The agents are rational maximizers.
16. The agents possess complete relevant information.

related forces, and concentric rings emerge.

It is remarkable that Thünen himself was not in the least confused about the method he applied. He was aware that his idealizations are false about the world, but he argued that they "are a necessary part of my argument, allowing me to establish the operation of a certain factor, a factor whose operation we see but dimly in reality, where it is in incessant conflict with others of its kind" (1966, pp. 3–4). By using his method (of isolation) "we may divest an acting force of all incidental conditions and everything accidental, and only in this way can we recognizing its role in producing the phenomena before us" (1910, p. 274; my translation).

4 Benchmark model idealization

Minimal models of the sort discussed above highlight the causal contribution of a causal unit to the emergence or existence of a phenomenon, using idealizations for neutralizing other causal factors. By contrast, using models as benchmarks highlights the way in which such idealized factors, once released from control, play their causal roles. This can mean several different things.

Thünen again
The Thünen model can again be used to illustrate. In the benchmark, the Thünen Mechanism, when acting alone, shapes the outcome pattern in the model, that of concentric rings (Figure 1, left). This is not sufficient for explaining the far less neat and regular patterns in the actual world (illustrated by an image of the larger Beijing area in Figure 1, right). The causal contribution of the Thünen Mechanism to shaping the actual pattern is modified by several other causes and conditions that can be mobilized in explaining the actual pattern by relaxing the idealizing assumptions that were used for neutralizing their impact in the simple model that isolated the Thünen Mechanism from everything else. The model is now de-isolated by de-idealization to let multiple other factors play their causal roles in shaping the outcome – such as rivers and mountains, several urban concentrations and their internal structure, various other heterogeneities

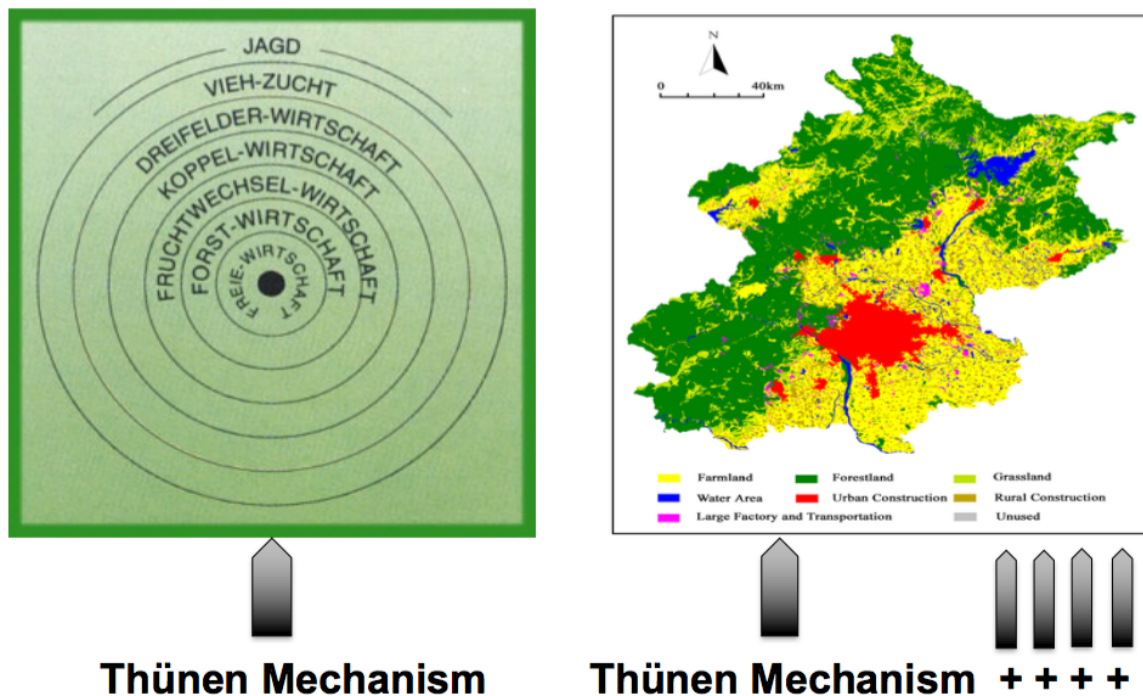as well as limitations in decision-makers' rationality.

Figure 1. Thünen model applied,
without and with relaxed idealizations

The whole empirical pattern is the *explanandum* here. In explaining it, one needs to invoke the Thünen mechanism *and* to relax at least some of the idealizations so as to enable invoking some further causes and conditions for explaining the irregular richness of the actual pattern. The idealizations of the simplest model play the role of early-step assumptions. On later steps, they are relaxed.

In a different strategy, one uses the Thünen model as a benchmark for identifying deviations from it in the actual world, and then sets out to explain those deviations by relaxing the idealizations of the model. The Thünen Mechanism is no more invoked in these explanations (Figure 2).
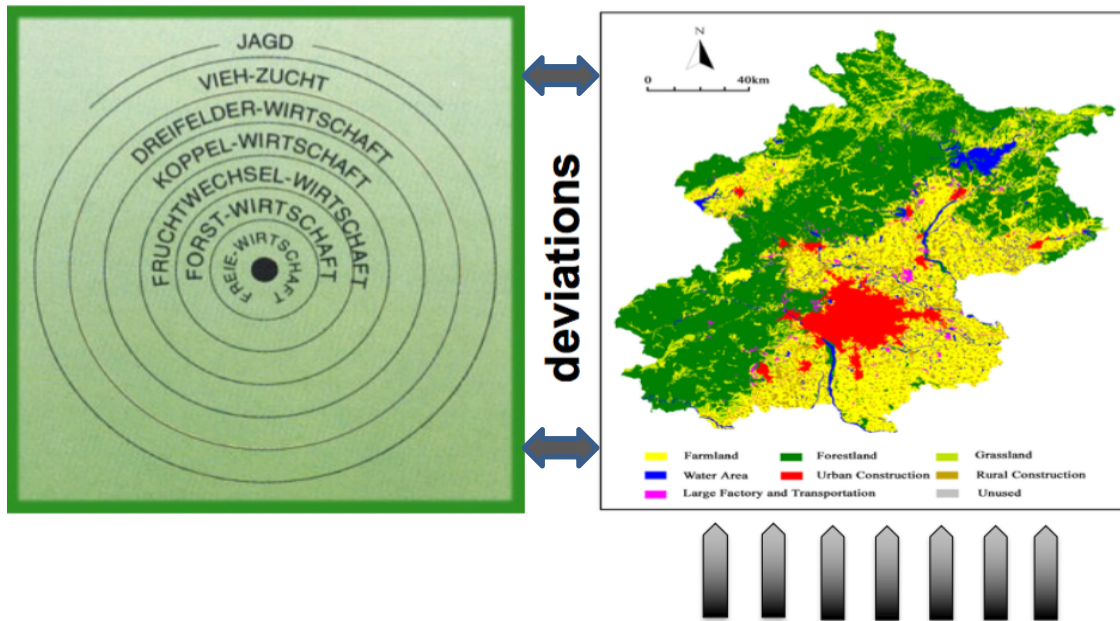
Figure 2. Thünen model used as
benchmark, deviations explained

Here *deviations* from the benchmark are the *explananda*, and they are explained in terms of factors that were idealized in the benchmark model. The idealizations are now relaxed depending on which theoretical enrichment is needed for explaining which deviation. This strategy provides a way of acquiring a sharper understanding of how each such subsidiary factor contributes to a feature in the outcome. This sharper view can be developed thanks to the contrast between the model where these factors are neutralized and the actual world where they are let to make their full impact on the outcome pattern. What is idealized away in the basic model is now brought to bear on the outcome, to contribute their share in the actual causal structure. The role of the Thünen Mechanism in shaping the actual pattern is not of interest in this strategy. Its role is decisive in the basic model, while the burden of explanation is now reallocated to the various other factors, such as rivers, mountains, heterogeneities in fertility and climate, extended urban structures, complexities in governance structures, bounded rationality, and so on.

We may put the difference between minimal model idealization and benchmark model idealization in terms of explanatory burden as follows. In the former, *what is included* in the basic model explains; in the latter, *what was excluded* from the basic model explains.

Max Weber and ideal types

Puzzlement reigns around the notion of ideal type. Max Weber himself was famously far from clear about what ideal types are. He did not provide one elaborate and unambiguous definition of the concept, and his examples of ideal types (bureaucracy, capitalism, *homo economicus*, etc.) appear to suggest different types of ideal types. The brief notes below are intended to highlight some selective aspects of the conundrum, leaving much space for further interpretive effort. The main focus will be on the benchmark function. Let me first briefly comment on the two components of 'ideal type', viz. 'type' and 'ideal'.

It seems ideal types in Weber's sense are *not types in the ordinary sense of having instances that are typical.* Types proper allow for exceptions, but they are abundantly exemplified. Weber (e.g. 1949, 90) is explicit that ideal types have no exemplifications, so they are not exemplified by nearly all actions or nearly all social patterns in the relevant set. We might want to say that Weber's use of 'type' is atypical.

As to 'ideal' in 'ideal type', there is no one clear-cut definition that we can find in Weber, but he mainly seems to think of it in terms of abstraction and accentuation (*Steigerung*). He says that concepts are "formed as ideal-types by abstracting and accentuating certain conceptually essential elements" and that "[e]very individual ideal type comprises both generic and ideal-typically constructed conceptual elements" (ibid., 100; see also 90). To get to an ideal type, one *first abstracts* common features from instances, *then accentuates* those abstracted elements.

This calls for some scrutiny. Abstracting common features of a number of instances can be taken to yield generic types that are exemplified by those instances; this would be in line with ordinary usages of 'type'. This feature of types (i.e. that they are exemplified by token instances) then vanishes at the next step. This step is *Steigerung*, analytical accentuation or intensification of certain properties, leading to something "like a utopia" that does not have instances.

I have two quick comments on this. First, we might conclude that Weber's thinking has *a tension* built into it, a tension between the two components of 'ideal type'. Ordinary types have instances, while Weberian ideal types don't, and they don't due to their ideality. One way of putting this is to say that *the ideality of ideal types destabilizes their typicality*. This is puzzling, of course, but I leave it here, without trying to offer any further alleviation. This would not seem to have any dramatic implications for our main concern, that of how ideal types function.

Second, now focusing on 'ideality' alone, it does not seem to be entirely clear that Weberian ideal types are idealizations, at least in our narrow sense. It seems Weber thinks ideality derives from *Steigerung* – accentuation, intensification, amplification. Now this can be a matter of just *exaggerating* a property or behavior, which would not yet constitute idealization in our sense. The exaggeration would have to be taken to extremes to get to idealization. On the other hand, there may be interesting differences between Weber's examples of ideal types; for example, the ideal type of *homo economicus* more obviously involves idealizations proper, while that of bureaucracy is a less obvious case. Anyway, we have another puzzlement here, but I leave this too unsettled.

Having made some remarks about the puzzling notion of ideal type itself, I next suggest what seems an obvious way in which ideal types are supposed to be *used*. This is their benchmark function, akin to the Thünen case illustrated in Figure 2.

Weber implies that ideal types *do not function in virtue of being types* that have typical instances: ideal type is not "fitted to serve as a scheme under which a real situation or action is to be *subsumed as one instance*" (ibid., 93; italics added). The explanatory functions of ideal types are rather *based on their ideality*: "It has the significance of a *purely ideal limiting concept with which the real situation or action is compared* and surveyed for the explication of certain of its significant components." (Ibid.; italics added.) In empirical research, the function of an ideal type is "the comparison with

empirical reality in order to establish its divergences and similarities, to describe them with the most unambiguous intelligible concepts, and to understand and explain them causally" (ibid., 43; see also 97).

The key point is that ideal types are supposed to be compared to actual situations, and it is not until these comparisons have been made and deviations have been identified that explanatory activity may start. For example, the idealizing assumptions of rational action in theoretical economics are indispensable means for explanatory purposes: "Only in this way is it possible to assess the causal significance of irrational factors as accounting for the deviations from this [ideal] type." (Weber 1947, 92) All errors and emotional influences can be seen as *Störungen*, disturbances that result in *deviations* from the ideal type. The crucial step is to identify these deviations as the proper *explananda*. Their explanation requires invoking perspectives other than economics, such as those of psychology and sociology; thus we here also have a recipe for division of disciplinary labor. Without the ideal type used as a benchmark, the disturbances and the deviations cannot be identified and assessed for their distinctive causal roles, nor can the relevant explanatory resources be identified and invoked. Again, in contrast to minimal model explanation, it is not what is included in the model that explains, but rather what is excluded from the ideal type that explains phenomena of interest.[7]

These observations should go at least some way towards removing puzzled confusions about how to use and assess Weberian ideal typical accounts. For example, Weber's ideal model of bureaucracy has been criticized unduly for misrepresenting actual bureaucracies by distorting or missing various important details in their structure, functioning, and context. This is a mistake about applicability conditions. As Lamond (1990) reminds us, Weber's model is not intended for being used in a microanalysis of organization, but rather in a comparative – cross-cultural and historical – analysis of the organization of society as a whole.

---

[7] Weber's thinking is related – both historically and comparatively – to Carl Menger's thinking about economics as an exact science that employs "exact types" and "exact laws" in explaining economic phenomena. See Mäki 1997.

5 How-possibly model idealization

Minimal models typically deal with how-actually cases. Even though such models do not cover the whole range of causes and conditions that jointly bring about the phenomenon to be explained, they are often used to suggest that the isolated cause actually contributes to its emergence or existence. The epistemic ambitions may also be more modest, such that only how-possibly explanations are pursued.

Economic model building often explores possible causal scenarios. One first recognizes a phenomenon, such as a pattern of behavior. One then proceeds abductively, asking and attempting to answer the question, "What mechanism(s) could have generated this pattern?" In answering the question, one proposes a model that accounts for the pattern by isolating a possible mechanism that could be causally responsible for the pattern. It is not established or even proposed as the mechanism that *actually* has produced the pattern (or has actually made a major contribution to its emergence). It may be just entertained as a member in a set of many possible mechanisms. Further empirical inquiry may then try to establish one (or some) of them as actually responsible for the phenomenon.

While there might be just one correct how-actually explanation (given the precise explanatory question that has been asked), there may be several correct how-possibly explanations (this is one justification for entertaining multiple models at the same time). For the latter to be judged correct, they must meet some constraints. Not just any imaginative logical possibility will do. The space of possibilities may be constrained by what is known about the nature of the real world, including natural, cognitive and social facts. This delimits the set of hypothetical scenarios that are considered worthy of further examination.

Historical examples in social science include Carl Menger's "conjectural history" of the emergence of money – as generally accepted medium of exchange – from barter (Menger

1892; see Aydinonat 2008). We know that such a medium exists, and Menger provides a simple how-possibly explanation for its emergence. It is based on an idealized model that excludes everything but separate self-interested economizing actors in the market, engaged in direct exchange of goods with varying saleability. In this idealized world, devoid of things such as some other institutions, government activity, and cultural meanings of money, money emerges as an unintended consequence of separate individuals' interactions. It is not necessarily argued that this is how money has historically emerged as a matter of actual fact, just that this is how money could possibly emerge. This is supposed to contribute to our understanding of the nature of money.

More recent examples include economic models of science (Mäki 2005b) and Thomas Schelling's models of segregation (Schelling 1978; Aydinonat 2007, 2008; Grüne-Yanoff 2009; Mäki 2009; Verreault-Julien 2019). Schelling's models have been much discussed recently, so I can briefly focus on just their function as how-possibly models.

Schelling's model is an agent-based model that traces processes that can generate segregated social structures. In one racial housing market version, the modeller starts with a random distribution of housing between blacks and whites. One then lets the residents follow rules governed by their (non-racist) preferences for their neighbors' skin color in their residential location decisions. Finally, racial segregation will emerge in the model city. This is presented as an unintended (and unwanted) consequence of interactive individual decisions.

Again, various idealizations contribute to cleaning the stage for the simple invisible-hand mechanism to produce the segregated outcome in isolation. It may be assumed that the city is spatially homogeneous and closed, that is, there are no neighbourhoods in the city and no moves to and from the city. Agents have uniform preferences and behave rationally. Agents make decisions about whether to move or stay based on just their preferences about their immediate neighbours, thus either have no other preferences or their other preferences have no influence on moving decisions. Agents move from one slot to another at no cost. Information about a slot left behind by an agent is immediately

available to all agents. There are no differences in housing prices and budget constraints. Socio-economic factors play no role. No external interference, such as by zoning authorities, shapes the process.

We know the world is not like that. Yet, the model might be informative about a possible way in which segregation emerges – or perhaps rather, about a possible contribution that something like this invisible-hand mechanism makes to the outcome. Whether it is informative about such a possibility is not dependent on whether it provides a complete story about how the segregated outcome actually comes about generally or in any particular case. Removing any possible confusions about these functions has implications for how Schelling's models are to be appropriately critically assessed.

It is remarkable that how-possibly models may have policy relevance. The Schelling case illuminates First there is the observed fact that racial segregation obtains. Second, we have a typical commonsense explanation: segregated structures obtain because people want to live next to their "own race". Third, an alternative how-possibly explanation is given by the model: the segregated structure is a possible unintended consequence of many alternative preferences, racist and non-racist alike. Fourth, a political conclusion transpires: rather than simply "let people have what they want" there may be possible grounds for a political intervention, by way of housing and zoning policies for example. Indeed, exploring possible causal scenarios with practical relevance is yet another possible function of idealizations.

6 Normative ideal idealization

We now come to our final example, one that involves normativity. In post-Rawlsian political philosophy, there is a methodological debate under way, criticizing or justifying what they call ideal theory in contrast to non-ideal theory. Some of the key concepts – including those of ideal and non-ideal theory – tend to be ambiguous (see e.g. Valentini 2012; Hamlin and Stemplowska 2012). Heightened puzzlement reigns. We need to ask questions about the concept of ideal theory itself, as well as about the functions of

idealization in ideal theory. The issues are vast and tangled, so only a few brief and comparative opening observations can be submitted.

In putting forth his "ideal theory" Rawls sets the stage for dealing with the problem of social justice by assuming a society that is well-ordered and self-contained[8], that its population consists of physically and mentally fully capable adults, and that everything takes place under "reasonably favourable" natural and historical conditions, such as in the absence of wars and natural disasters. Importantly, the agents are assumed to strictly comply with his principles of justice. Ideal theory "assumes strict compliance and works out the principles that characterize a well-ordered society under favorable circumstances" (Rawls 1971, 245).

I have two observations to make about full or strict compliance, the key idealization to assume away any motivational, cognitive and behavioral deficiencies, required for isolating the fundamental principles of justice. First, regarding the contents of the idealization, saying that such weaknesses must be excluded when they are "beyond reasonable extent" (Valentini 2009, 339) is to pass a negligibility judgement, an altered version of the idealization of full compliance. This is the claim that any such actual weakness *not* beyond reasonable extent is negligible (given what the ideal theory is used for). This means that 'strict compliance' does not have to be taken in a strictly literal manner in all applications. The same observation applies to the assumption of "reasonably favourable circumstances" which is an altered version of the idealization of "perfectly favourable" conditions.

Second, one may be puzzled about the *function* of the idealization of full compliance. An obvious response is to suggest that the function is that of "experimental control" for the varying less-than-perfect degrees of compliance in actual behavior so as to isolate the "pure" effects of society's ordering principles. This is recognized by John Simmons:

---

[8] In the theory of domestic justice, the relevant community unit is assumed to be closed to cross-border interactions, just as are our other examples, those of a closed economy, Thünen's isolated state, and Schelling's segregated cities.

> "if we compare the operation of societies ordered by competing principles of justice while assuming strict compliance with those principles, the different effects we observe can reasonably be taken to be wholly the responsibility of the different ordering principles themselves. So our comparison turns out to be quite strictly a comparison only of the principles of justice." (Simmons 2010, 8)

This can be put as follows. By neutralizing the effects of any varying and hard-to-measure degrees of non-full compliance, we can focus on examining the effects of the ordering principles without the interference of failures to fully comply. This manifests an "experimental" use of minimal-model idealization. On the other hand, if this procedure is not applied, and the degree of compliance below perfection is not neutralised for its effects, we lose the prospect of isolating the causal effects of other factors as they are intermingled with those of imperfect compliance. On top of possibly unfavourable conditions, these other factors include other valuable goals such as peace, security, order, legitimacy, democracy. An idealized theory of justice identifies and examines principles of justice in isolation from all other normative principles.[9]

The above does not exhaust the functions of ideal theory. Normativity is also involved. Ideal theory provides

> "a conception of a just society that we are to achieve if we can. Existing institutions are to be judged in the light of this conception" (Rawls 1971, 246).

This evokes two normative functions of the ideal. One is a *prescriptive function*: we should work toward the ideal. "We are to achieve it if we can" prompts a couple of observations. First, the ideal is a valuable goal to be pursued and achieved as an outcome of those pursuits. This contrasts with outcomes of idealized situations such as

---

[9] This gives rise to issues related to possible trade-offs between the various normative goals as well as to possible second-best arguments against simplistic advice for proceeding towards the goal of justice.

segregated housing markets (Schelling) and concentric rings in land use (Thünen) that are not in themselves valuable goals to be pursued. Second, "if we can" suggests that 'ought' implies 'can' and that feasibility issues will be important. Indeed, feasibility has been an important concern in the debates over just society.

The other normative function is *evaluative*: actual situations are to be judged against the ideal. This suggests that the ideal of a perfectly just society also has the function as a benchmark, a standard for identifying and estimating deviations from the ideal in the actual world. This links with our earlier discussion on the benchmark idealization, but again there is an important difference. Here the benchmark is normative, and so the deviations do not primarily constitute *explananda* for social scientific explanations; they rather call for action that would bridge the gap between the ideal and the actual.[10] This is where non-ideal theory is supposed to come to the fore.

Non-ideal theory exhibits characteristics of later-step models in scientific modelling, those constructed by relaxing one or more of the idealizations of ideal theory to explain further features in the *explanandum* phenomenon (in the vein of Figure 1 above). Again, here the relationship between ideal and non-ideal theory is more complex than this. Ideal theory is about ends, non-ideal theory is about the means or the paths, or the distinction is between an end-state theory and a transition theory, the latter examining how we get from here to there. One of the issues in the debate is about whether the ultimate goal of perfectly just society set by ideal theory is required at all for us to have an "action-guiding" idea of the right direction towards a more just society – whether we need to know the best to know what is better.

These comparative observations have illuminated some of the similarities and differences between idealization in scientific modelling and in political philosophy. Admittedly we are far from having removed all the puzzlements and confusions – but

---

[10] Yet I believe these two (deviations as *explananda* for scientific explanation and as calls for action) had better be combined. Proposals for actions needed for getting from the actual to the ideal situation should be based on explanatory information about the gap between the two.

luckily that particular ideal end-state was not our goal, so it is enough if we have managed to step on the non-ideal path of transition towards it![11]

7 Conclusion

Idealization is a ubiquitous, powerful, and risky cognitive instrument. That there are suspicions and confusions about idealizations in general or about particular idealizations is unsurprising given that they appear to conflict with known facts and that it is often unclear what their point is.

I have proposed alleviating these puzzlements by drawing attention to the variety of claims that idealizations can be used to make (about negligibility, applicability, early-step status, etc.) as well as to the variety of functions that idealizations can serve (such as minimal, benchmark, and how-possibly modelling). I am not pretending to have identified a full list of possibilities, but the main point is that any particular idealization is to be assessed against some such decent function, not solely in terms of its truthlikeness or the like. This should reduce puzzlement and improve the quality of critical conversation on idealizations. Particular idealizations may go astray, including those that I have used as illustrations in this paper. But an idealization with this or that content fails or succeeds in relation to this or that function, therefore we need to get those contents and functions clear to be able to determine whether an idealization does its job or fails to do so.

---

[11] There are further puzzlements and confusions in the conversations on ideal and non-ideal theory in political philosophy, including those around the contrastive uses of the notions of 'abstraction' and 'idealization' since Onora O'Neill's early contributions (O'Neill 1987), but I leave their investigation for another occasion.

References

Aydinonat, N. Emrah (2007) "Models, conjectures and exploration: an analysis of Schelling's checkerboard model of residential segregation" *Journal of Economic Methodology*, 14(4): 429 – 454.

Aydinonat, Emrah (2008) *The Invisible Hand in Economics*. London: Routledge.

Blau, Peter (1956) *Bureaucracy in Modern Society*. New York: Random House.

Cartwright, Nancy (1981) *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Friedman, Milton (1953) 'The methodology of positive economics', in *Essays in Positive Economics*, Chicago: Chicago University Press pp. 1–43. Reprinted in Mäki 2009.

Grüne-Yanoff, Till (2009) "Learning from minimal economic models" *Erkenntnis* 70(1): 81-99.

Hamlin, Alan; Zofia Stemplowska (2012) "Theory, ideal theory and the theory of ideals" *Political Studies Review* 10, 48-62.

Hindriks, Frank (2006) "Tractability Assumptions and the Musgrave-Mäki Typology" *Journal of Economic Methodology* 13: 401–423.

Jones, Martin (2005) "Idealization and abstraction: A framework" *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 86: 173–217.

Lamond, David A. (1990) "The irrational use of Weber's ideal types" *Australian Journal of Public Administration* 49 (4): 464-473.

Mäki, Uskali (1992a) "On the method of isolation in economics", *Poznan Studies in the Philosophy of the Sciences and the Humanities,* 26, 319-354.

Mäki, Uskali (1992b) "Friedman and realism", *Research in the History of Economic Thought and Methodology*, 10: 171-195.

Mäki, Uskali (1997) "Universals and the *Methodenstreit*: A reexamination of Carl Menger's conception of economics as an exact science" *Studies in History and Philosophy of Science,* 28, 475-495.

Mäki, Uskali (2000) "Kinds of assumptions and their truth: Shaking an untwisted F-twist" *Kyklos,* 53(3), 303-322.

Mäki, Uskali (2005a) "Models are experiments, experiments are models" *Journal of Economic Methodology,* 12, 303-315.

Mäki, Uskali (2005b) "Economic epistemology: Hopes and horrors" *Episteme. A Journal of Social Epistemology,* 1(3): 211-220.

Mäki, Uskali (2009) "Unrealistic assumptions and unnecessary confusions: Rereading and rewriting F53 as a realist statement", in *The Methodology of Positive Economics. Reflections on the Milton Friedman Legacy,* ed. U. Mäki*.* Cambridge University Press. Pp. 90-116.

Mäki, Uskali (2011) "Models and the locus of their truth", *Synthese* 180, 47-63.

Mäki, Uskali (2012) "The truth of false idealizations in modeling" in *Models, Simulations*, *and Representations*. Ed. P. Humphreys and C. Imbert. London: Routledge. Pp. 216-233.

Mäki, Uskali (2013) "Contested modeling: The case of economics" in *Models, Simulations, and the Reduction of Complexity*. Ed. U. Gähde, S. Hartmann and J.H. Wolf. Berlin / New York: Walter de Gruyter. Pp. 87-106.

Menger, Carl (1892) "On the origin of money", *Economic Journal*, 2, 239-255.

Mill, John Stuart (1844) "On the definition of political economy; and on the method of investigation proper to it" in *Essays on Some Unsettled Questions of Political Economy.* London: Longmans*.* Pp. 120-164.

Musgrave, Alan (1981) '"Unreal assumptions" in economic theory: The F-twist untwisted', *Kyklos*, 34: 377–87.

Nowak, Leszek (1980) *The structure of idealization*. Dordrecht: Reidel.

O'Neill, Onora (1987) "Abstraction, idealization and ideology in ethics" *Royal Institute of Philosophy Supplements*, Volume 22, 55-69.

Potochnik, Angela (2017) *Idealization and the Aims of Science.* Chicago: Chicago university Press.

Rawls, John (1971) *A Theory of Justice*. Cambridge MA: Harvard University Press.

Schelling, Thomas C. (1978). *Micromotives and macrobehavior*. New York: Norton.

Simmons, A. John (2010) "Ideal and non-ideal theory" *Philosophy & Public Affairs* 38 (1): 5-36.

von Thünen, J. H. (1910). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Jena: Verlag von Gustav Fischer.

von Thünen, J. H. (1966). *Von Thünen's isolated state* (P. Hall, Ed., C. M. Wartenberg, Transl.). Oxford: Pergamon.

Valentini, Laura (2009) "On the Apparent Paradox of Ideal Theory" *The Journal of Political Philosophy,* 17(3): 332–355

Valentini, Laura (2012) "Ideal vs Non-ideal Theory: A Conceptual map" *Philosophy Compass*, 7/9, 654-664.

Verreault-Julien, Philippe (2019) "How could models possibly provide how-possibly explanations?" *Studies in History and Philosophy of Science* 73, 22-33.

Weber, Max (1947) *The Theory of Social and Economic Organization*. Edited and translated by A.M. Henderson and T. Parsons. New York: Free Press.

Weber, Max (1949) *Max Weber on the Methodology of the Social Sciences*. Ed. E.A. Shils and H.A. Finch. Glencoe, IL: Free Press.

Weisberg, Michael (2007) "Three kinds of idealization" *Journal of Philosophy* 104 (12): 639-659.