Journal of Philosophical Logic (2020) 49:1159–1183
https://doi.org/10.1007/s10992-020-09552-x

Predicate Change

A Study on the Conservativity of Conceptual Change

Corina Strößner[1]

Received: 3 May 2019 / Accepted: 14 February 2020 / Published online: 26 March 2020

Abstract
Like belief revision, conceptual change has rational aspects. The paper discusses this for predicate change. We determine the meaning of predicates by a set of imaginable instances, i.e., conceptually consistent entities that fall under the predicate. Predicate change is then an alteration of which possible entities are instances of a concept. The recent exclusion of Pluto from the category of planets is an example of such a predicate change. In order to discuss predicate change, we define a monadic predicate logic with three different kinds of lawful belief: analytic laws, which hold for all possible instances; doxastic laws, which hold for the most plausible instances; and typicality laws, which hold for typical instances. We introduce predicate changing operations that alter the analytic laws of the language and show that the expressive power is not affected by the predicate change. One can translate the new laws into old laws and vice versa. Moreover, we discuss rational restrictions of predicate change. These limit its possible influence on doxastic and typicality laws. Based on the results, we argue that predicate change can be quite conservative and sometimes even hardly recognisable.

Keywords Conceptual change · Dynamic epistemic logic · Belief revision · Analyticity · Typicality · Conditional logic

1 From Doxastic Dynamics to Conceptual Change

The last decades, starting at latest with the work of [1], have seen flourishing research in the study of epistemic and doxastic dynamics. Dynamic epistemic logics (DEL),

✉ Corina Strößner
stroessner@phil.hhu.de

[1] Duesseldorf Center for Logic and Philosophy of Science, Heinrich Heine University Düsseldorf, Building 24.52 Universitätsstraße 1, 40225 Düsseldorf, Germany

where modal operators can express what holds after a change of the model, are a currently intensively investigated branch of this research.

The paradigmatic example of DEL is PAL, the logic of public announcement. The core idea of PAL is to model the effects of the public announcement of a statement P on the knowledge of (groups of) agents. This is realised by a model change, namely a reduction of the epistemic possibilities to worlds that comply with the announced P. The most crucial effects concern the *common knowledge* of groups, i.e., what every member of the community knows, and what is known to be known by everyone, and known to be known to be known, and so on [18].[1] Though public announcement alters the truth-value of epistemic expressions, PAL is not primarily a logic of *meaning* change but of *epistemic* change. Non-epistemic formulae cannot alter their meaning, i.e., the set of worlds in which they are true. This restriction was explicitly formulated in the foundational work on PAL by [3]:

> All of our work rests on the assumption that our actions do not change facts, so the atomic propositions true at (w, a) [after the change] are going to be the same as those true at w in W [3, 46].[2]

Whether one knows or believes in a world that the morning star is a planet can be changed by public announcement, but nothing can change whether the morning star *is* a planet. This is seemingly beyond our power.

While the status of Venus as a planet was indeed never doubted, the recent finding of Eris, a dwarf planet bigger than Pluto, triggered a change in the concept of planets that eventually lead to the still controversial exclusion of Pluto from the category of planets. This was only the most recent of many changes the concept PLANET underwent in its long history.[3] This is exactly the kind of dynamics we deal with in this paper. Arguably, this change is neither an ontic (real world) change nor a doxastic one, but a conceptual change. However, since conceptual change is directly related to beliefs, it is part of a larger theory of rational doxastic dynamics.

How do we tell whether a change is conceptual? When do we merely learn new important facts about a concept and when do we truly change it? For one thing, other than any kind of belief enrichment or revision, conceptual change can alter the way we discriminate between different possibilities. The distinction between heat and temperature, the rising of the concept of mass, or the introduction of complex numbers give a new structure of epistemic possibilities. Such severe changes, dubbed "conceptual revolution" by [30], are often associated with different thought styles or incommensurability (cf. [12, 16]). Within this paper, however, we focus on a less drastic variant of conceptual change, namely predicate change.

[1] We do not repeat the many results and details of PAL, which can be found in the extensive literature on DEL. The standard introduction to the field is [11]. Van Benthem [5] is a comprehensive study that illustrates and uses DEL methodology in a formal inquiry of intelligent interaction. More recently, the Stanford encyclopaedia of philosophy included a comprehensive entry on the subject [2].

[2] The basic assumption of DEL has meanwhile been challenged by [6] who propose LCC, a logic that considers changes on V as well. Such operations are considered to capture "real world change" by [5, 85].

[3] We use capitalised words to refer to concepts.

Predicate change alters the set of *conceivable* objects that count as instances of the concept, including existing instances as well as merely imaginable ones. Discovering a new planet, e.g., Neptune in 1846, is no conceptual change in this sense. In the conceptual system of that time, any Neptune like entity would have been categorised as a planet. What was learnt is that such category member truly exists, i.e., that it is not only an imaginable but a real instance of the concept. The quite recent exclusion of Pluto, on the other hand, was clearly a consequence of the redefinition of PLANET and thus a predicate change.

The following two sections define two monadic predicate logics with three different kinds of operators that stand for decreasingly strict beliefs: analytic laws, doxastic laws, and typicality laws. In the fourth section, we will apply the main idea from DEL, namely model changing operations, to represent predicate change. Though the operations affect the analytic laws (meanings) of the language, we show that the model change entails no incommensurablity: one can translate statements from the old to the new system. In the remaining sections, we discuss predicate changes with respect to other laws by raising two questions: First, how can doxastic or typicality laws motivate a conceptual change? Second, how can predicate change be evaluated with respect to the consequences it has for these other laws? The final conclusion gives a sketch of how our results can be applied to two central subjects of philosophy, namely the understanding of conceptual change and the role of analytic laws.

2 AD: Predicate Logic of Analytic and Doxastic Laws

This section will define a monadic logic of analytic and doxastic laws, henceforth called AD, where the domain includes all imaginable objects. Technically, AD resembles a propositional modal logic with unary predicates instead of propositions and possible objects instead of possible worlds.[4] A plausibility relation orders these objects:

Definition 1 (Model of AD) A model $\mathbf{M}$ of AD is a triple $\langle O, I, \geqslant \rangle$ where O is a finite set of (conceptually) possible objects, I is an interpretation that assigns subsets of O to predicates, and the plausibility order $\geqslant$ is a totally connected, reflexive, and transitive plausibility relation.

AD contains atomic predicates as well as complex predicates, which are formed by set-theoretic notions:

Definition 2 (Predicates in AD) Only the following strings are predicates of AD:

1. S, P, Q with or without natural numbers as subscript are atomic predicates of AD.
2. If Φ and Ψ are predicates of AD, then

[4] Readings of propositional logic as unary predicate logic are not new in the field (cf. [31]).

(a) $-\Phi$,
(b) $\Phi \cup \Psi$,
(c) and $\Phi \cap \Psi$

are predicates of AD.

The interpretation function I ranges over all predicates and assigns possible objects to them:

$I(\Phi) \subseteq O$, such that

(a) for $I(-\Phi) = O/I(\Phi)$,
(b) for $I(\Phi \cap \Psi) = I(\Phi) \cap I(\Psi)$,
(c) and for $I(\Phi \cup \Psi) = I(\Phi) \cup I(\Psi)$.

The elements of O are conceivable, i.e., conceptually possible objects. While the notion of possible worlds is highly prominent in philosophical (i.e., modal) logic – it is its most central notion – possible or conceivable objects are less familiar, though such notions are found early in the development modal logic [17]. Conceivable objects can be matched to a maximal consistent description in terms of predicates. However, they cannot be identified with a set of predicates since, if the meaning of the language will change, the description of these objects will change accordingly. Moreover, one should not presuppose that every objects has its own unique description in a particular state of the conceptual development. In principle, conceptual dynamics might lead to states in which previously indistinguishable objects become distinguishable and vice versa. Within the present investigation, however, this will not happen.

Examples of possible objects include quite common things like green, round apples or sweet strawberries, but also flying dogs, talking penguins, or unicorns. Like possible worlds, the objects O are maximally consistent. For every $o \in O$ and every atomic predicate Φ with its complement $-\Phi$ it holds that either $o \in I(\Phi)$ or $o \in I(-\Phi)$. We call $I(\Phi)$ the category of Φ. Technically, the interpretation function I resembles the extension of a predicate in FOL (First-order predicate logic), but categories include *possible* objects and are thus intensional. They are what [17, 238] calls *comprehension*: "the classification of all consistently thinkable things to which the term would correctly apply". For example, the category corresponding to ANIMAL WITH A HEART is not identical to the one for ANIMAL WITH A LIVER since there are arguably conceptually possible animals that have only one of these organs, though we do not believe that they exist.

The relation $\geqslant$ encodes the plausibility of possible objects. $x \geqslant y$ means that, according to one's beliefs, x is at least as plausible (to exist) as y. For any two possible objects $x, y \in O$, one of the following possibilities holds:

– x is as plausible as y: $x \triangleq y := x \geqslant y \;\&\; y \geqslant x$
– x is more plausible than y: $x > y := x \geqslant y \;\&\; y \not\geqslant x$
– x is less plausible than y: $x < y := x \not\geqslant y \;\&\; y \geqslant x$

Note that we do not allow that two objects are incomparable: at least one, $x \geqslant y$ or $y \geqslant x$, must hold, because $\geqslant$ is totally connected.

Two different kinds of laws are the core of AD: analytic laws ($\mathcal{A}$) and doxastic laws ($\mathcal{D}$). The $\mathcal{A}$ laws concern all possible instances of a concept, i.e., their intension. This is why we call them analytic laws.

Definition 3 ($\mathcal{A}$ laws) $\mathcal{A}$ laws (analytic laws) are sentences of AD with the following syntax and semantics:

If Φ and Ψ are predicates of AD, then $\mathcal{A}\Phi\Psi$ is a sentence of AD.
$\mathbf{M} \models \mathcal{A}\Phi\Psi$ iff $I(\Phi) \subseteq I(\Psi)$.

The idea behind the semantics of $\mathcal{A}$ is that an exception is not imaginable, i.e., *conceptually* impossible. “All mammals are animals” is such an analytically true statement. It is impossible to describe a mammal that is not an animal. Analyticity restricts what is conceptually conceivable. “All mammals are born on earth” is not an analytic law. Nothing is conceptually wrong about the assumption that some mammals travelled to another planet 3000 years ago and spread there. Analytic laws are thus extremely strict. In evaluating them, we need even consider highly far-fetched and implausible exceptions.

$\mathcal{D}$ laws are restricted to members that have a more privileged role in the belief systems. This is why we call them doxastic laws. For modelling these laws, we need to take the plausibility relation $\geqslant$ into account. Roughly speaking, it creates a system of plausibility spheres. The top level includes all objects that we believe to exist or which we at least do not disbelieve to exist. Many universal statements only deal with such maximally plausible entities. For example “All animals with a heart have a liver” is a true law, but not for conceptual reasons but because *empirical* knowledge tells us that animals with a heart but without a liver do not exist.

The most plausible objects of a model $\mathbf{M}$ can be designated as follows: $P_{\mathbf{M}}(O) = \{x \in O | \neg\exists y(y \in O \wedge x < y)\}$. They comply with our current beliefs, our experiences and the laws of nature (white dogs with four legs, yellow apples and so on). This excludes entities that do not comply with our currently hold laws of nature (flying apples, immortal animals) as well as kinds of objects that are not violating laws of nature but about the existence of which we have no confirming observations (square apples, dogs with a horn). In the plausibility ordering, the law-violating entities are usually on a much lower level than kinds of entities which were never observed but which comply with beliefs about general laws of nature. There are concepts, for example, UNICORN, that do not have any instances we believe to exist. However, one can still distinguish instances that are more plausible (horse with a mutation for growing a horn) and instances that are bizarre (horse with magic abilities). The former is still compatible with most laws of biology while the latter is purely fictitious. Doxastic laws are thus not only discriminating between entities we believe to exist and those we disbelieve to exist, but allow for finer graduations among the entities we do not believe to exist.

There are concepts such as FURRED ANIMALS WITH A BEAK, which were not thought to have plausible instances but it turned out that they truly exist (platypus). Though we are mostly concerned with the maximally plausible objects of our conceptual system, the plausibility ordering with its degrees of implausibility and entrenchment influences classifications. For example, the (false) assumption that all animals with mammary glands are viviparous was highly influential in the classification of platypus. The biological system of Europeans before 1800 strongly hold that any suckling animal is viviparous and this was extended to animals they did not believe to exist, including an animal that appeared to be a duck-billed beaver. One might object that biologists only formed beliefs about platypus when they realised that such an animal exists. This is true in the sense that they did not have any (consciously) performed beliefs about the animal before they encountered it. However, the belief that suckling creatures are viviparous must have come from somewhere, namely from an already build doxastic preference. In case of platypus this previously formed doxastic priorities made European scientists even ignore actual evidence, namely testimonies from native Australians stating that platypus lays eggs (cf. [13]). This kind of entrenchment is only captured by taking the whole plausibility relation into account, which requires to formulate the notion of plausible members relative to categories:

Definition 4 (Plausible members) For a model $\mathbf{M} = \langle O, \geqslant, I \rangle$, $P_{\mathbf{M}}(\Phi)$ is the set of (maximally) plausible members in the category of Φ. $P_{\mathbf{M}}(\Phi) = \{x \in \mathbf{I}(\Phi) | \neg \exists y (y \in I(\Phi) \wedge x < y)\}$.

On this basis, we introduce the doxastic laws of AD. They express what is believed about the plausible members of a category:

Definition 5 ($\mathcal{D}$ laws) $\mathcal{D}$ laws (doxastic laws) are sentences of AD with the following syntax and semantics:

If Φ and Ψ are predicates of AD, then $\mathcal{D}\Phi\Psi$ is a sentence of AD.
$\mathbf{M} \models \mathcal{D}\Phi\Psi$ iff $P_{\mathbf{M}}(\Phi) \subseteq I(\Psi)$.

$\mathcal{D}$ laws can be revised by changing the plausibility relation $\geqslant$. The research on belief revision lead to intensive investigations on this subject. Rott [26] names no less than twenty-seven possible options to update (doxastic) priorities. One of them, the lexicographic update, has also been introduced in DEL by [4, 13]. His results are transferable to the logic AD, allowing to model changes of doxastic laws as changes of $\geqslant$, without any changes in the $\mathcal{A}$ laws, i.e., the conceptual basis of the language. By this, we do not deny that drastic doxastic changes have consequences for concepts, as the above-mentioned debate about platypus shows. Within this investigation, however, we will focus on the dynamics of analytic laws. The rather complex interaction with belief change is an issue for further research.

Finally, we introduce the classical connectives from propositional logic to complete the language AD.

Definition 6 (Connectives) The connectives of classical propositional logic can be used to combine $\mathcal{D}$ and $\mathcal{A}$ laws:

If ϕ and ψ are sentences of AD, then $\neg\phi$, $\phi \wedge \psi$, and $\phi \vee \psi$ are sentences of AD.

$\mathbf{M} \models \neg\phi$ iff $\mathbf{M} \not\models \phi$.
$\mathbf{M} \models \phi \wedge \psi$ iff $\mathbf{M} \models \phi$ and $\mathbf{M} \models \psi$.
$\mathbf{M} \models \phi \vee \psi$ iff $\mathbf{M} \models \phi$ or $\mathbf{M} \models \psi$.

Though they combine predicates, the laws of AD resemble conditional operators in modal logic. $\mathcal{A}$ is a strict conditional operator, while $\mathcal{D}$ is a variably strict conditional in the tradition of [19]. Note that $\mathcal{A}$ laws could also be expressed by a unary operator, defined as $\boxed{\mathcal{A}}\Psi =: \mathcal{A}(\Phi \cup -\Phi)\Psi$, i.e. as necessary truth for all entities. $\mathcal{A}\Phi\Psi$ is equivalent to $\boxed{\mathcal{A}} - (\Phi \cup -\Psi)$. For $\mathcal{D}$ laws, this is not possible. They are intrinsically binary and neither reducible to nor definable from other unary operations. Moreover, they are non-monotonic: $\mathcal{D}\Phi\Psi$ does not imply $\mathcal{D}(\Phi \cap \Xi)\Psi$. Those Φ that are Ξ could be less plausible than Φ that are not Ξ. Then the set $P_{\mathbf{M}}(\Phi \cap \Xi)$ is not a subset of $P_{\mathbf{M}}(\Phi)$ and can give rise to completely different doxastic laws for $\Phi \cap \Xi$. Only the following restricted form of monotonicity, dubbed rational monotony by [15], is valid:

$$\text{If } \mathbf{M} \models \mathcal{D}\Phi\Psi \text{ and } \mathbf{M} \models \neg\mathcal{D}\Phi - \Xi, \text{ then } \mathbf{M} \models \mathcal{D}(\Phi \cap \Xi)\Psi$$

If the premises are true, then $P_{\mathbf{M}}(\Phi \cap \Xi)$ is a subset of $P_{\mathbf{M}}(\Phi)$, namely $P_{\mathbf{M}}(\Phi) \cap \Xi$. Only if no member of $P_{\mathbf{M}}(\Phi)$ was Ξ, can different members of Φ dominate $\Phi \cap \Xi$. However, this is excluded by $\neg\mathcal{D}\Phi - \Xi$.[5]

Both laws, $\mathcal{A}\Phi\Psi$ and $\mathcal{D}\Phi\Psi$, correspond to the natural language statement "All Φ are Ψ". $\mathcal{A}\Phi\Psi$ means that an object Φ, which is not Ψ, is not conceivable (conceptual universality) while $\mathcal{D}\Phi\Psi$ captures the disbelieve in plausible Φ that are not Ψ (doxastic universality). In order to capture beliefs about what is normally true about Φ, we need yet another variant of lawful generalisations, namely typicality laws.

3 ADT: Typicality Laws

The relation $\geqslant$ is about general plausibility and thus not specific to a category. This is why it cannot be used to capture that penguins are exceptional *birds*, though they are not exceptional *animals*. What makes penguins a poor instance of BIRD is not a lack of plausibility but a low typicality.

In order to formulate typicality laws, we need a further operator $\mathcal{T}$ and a richer model. In particular, we want to assume that some concepts have their own specific (typicality) ranking.[6] We create these typicality orderings by adding two further components to the model: central category members – so-called prototypes – and

[5]Note that this inference pattern is only valid because we assumed that plausibility is totally connected, which creates a sphere system of decreasing plausibility.

[6]Note that we do not assume that every predicate has a *proper* typicality ranking. Concepts in the formal sciences (number, operators, etc.) usually lack a typicality structure. Schurz [28] argues that typicality is common in the social and biological domain because they were formed by evolutionary processes. This includes most of the concepts that are central to common sense reasoning.

comparative similarity to relate other entities to the central member. The enrichment gives us the following model:

Definition 7 (Model of ADT) A model **M** of ADT is a quintuple $\langle O, I, \geqslant, \succsim_{-}, Pt\rangle$ where O, I, and $\geqslant$ are as in AD models, and $\succsim_{-}$ is a ternary, weakly centred comparative similarity relation, and Pt is a partial function that assigns a prototype to some atomic predicates.

The relation $\succsim_{-}$ compares possible objects with respect to their similarity. $x \succsim_{o} y$ is read as "x is at least as similar to o as y". For every o, $\succsim_{o}$ is transitive and reflexive, as one would expect from a comparative relation. Moreover, $o \succsim_{o} x$ holds for every x since $\succsim_{-}$ is weakly centered. Every object is at least as similar to itself as any other object. It is, however, not necessarily *more* similar to itself then other entities. We allow that different objects are similar enough to o to qualify as equally similar to o as o to itself, i.e., strong centering is rejected. Finally, the partial prototype function Pt can be used to designate a central instance of an atomic predicate Φ. By means of the prototype $Pt(\Phi)$, and the similarity relation $\succsim_{-}$, we define a typicality ordering for Φ in the following way:

Definition 8 (Typicality) $x \succeq_{\Phi} y$ if and only if $x \succsim_{Pt(\Phi)} y$.

Thus, we judge whether a category member is a typical instance of a predicate by its comparative similarity to the prototype. This formal definition borrows a central idea from the prototype theory of concepts, namely that categories are internally structured by similarity to a prototype (cf. [14, 24, 25]). Nevertheless, ADT should be understood as a formal language and not as a psychological model.

Since we assumed that similarity is weakly centred, the prototype is always a typical member of the category: for all possible objects x it holds that $Pt(\Phi) \succeq_{\Phi} x$. However, the prototype doesn't need to dominate the typicality structure. There can be objects x such that $x \succeq_{\Phi} Pt(\Phi)$: they are just as typical as the prototype. The prototype is usually not the only typical category member.

Note that typicality is not totally connected. Two entities can be incomparable regarding the similarity to $Pt(\Phi)$ – neither $x \succsim_{Pt(\Phi)} y$ nor $y \succsim_{Pt(\Phi)} x$ is true – and so they are not related to each other in the typicality structure of Φ. The intuitive appeal of not requiring total connectedness is that we are not forced to compare the typicality of entities that are exceptional in very different respects. For example, penguins and peacocks are extraordinary birds, but it might be hard to compare them with respect to their typicality. The typicality structure also applies outside of the respective category. For example, a bat is not a bird but is more bird-typical than a mouse.

There are different possibilities to relate two objects in the typicality ordering:

– Incomparability
  $x \nsim_{\Phi} y := x \nsucceq_{\Phi} y \;\&\; y \nsucceq_{\Phi} x$
– Equal typicality
  $x \simeq_{\Phi} y := x \succeq_{\Phi} y \;\&\; y \succeq_{\Phi} x$

– Strict domination

$x \succ_\Phi y := x \succeq_\Phi y \; \& \; y \not\succeq_\Phi x$

$x \prec_\Phi y := x \not\succeq_\Phi y \; \& \; y \succeq_\Phi x$

If an atomic predicate has no prototype, then it has no proper typicality ordering and all entities are incomparable with respect to their typicality. The same holds for complex predicates: they have no prototype that can be used to determine the typicality structure.

In contrast to many proponents of Prototype theory (e.g., [14, 24]), we do not assume that degrees of typicality predict membership. The membership in Φ is fully fixed by $I(\Phi)$ and our model allows that objects that are quite similar to $Pt(\Phi)$, typical in this sense, don't actually belong to the category Φ. Only the prototype itself is guaranteed to be a member of the category.[7]

Our model gives us two options to designate a set of typical members. First, there are *prototypical* members of a category, namely members that are as close to the prototype as the prototype to itself. These prototypical members, however, can be quite implausible. Second, there are *plausible typical* category members. These are the most typical members among the most plausible members. From now on, we will just call them typical members. They are formally defined on the basis of plausible category members as follows:

Definition 9 (Typical members) For a model $\mathbf{M} = \langle O, I, \geqslant, \succsim_{-}, Pt \rangle$, $T_{\mathbf{M}}(\Phi)$ is the set of typical members. $T_{\mathbf{M}}(\Phi) = \{x \in P_{\mathbf{M}}(\Phi) | \neg \exists y (y \in P_{\mathbf{M}}(\Phi) \wedge y \succ_\Phi x)\}$

Note that the prototype $Pt(\Phi)$ is not necessarily an element of $T_{\mathbf{M}}(\Phi)$. It can be a very implausible entity. For example, a prototype bird could be a mixture of different bird species, but this mixture is an abstraction and does not exist. In determining the set of typical members, plausibility is of primary importance.

Typicality laws are defined as universal laws among the typical members:

Definition 10 (ADT and $\mathcal{T}$ laws) The logic ADT is defined as an extension of AD that includes typicality laws:

All predicates and formulae of AD are predicates and formulae of ADT.
If Φ and Ψ are predicates of ADT, then $\mathcal{T}\Phi\Psi$ is a sentence of ADT.
$\mathbf{M} \models \mathcal{T}\Phi\Psi$ iff $T_{\mathbf{M}}(\Phi) \subseteq I(\Psi)$.

That means, by $\mathcal{T}\Phi\Psi$, read as "Typical Φ are Ψ", it is claimed that Ψ is a universal property of typical Φ. This usage of the term "typicality law" may not do justice to the many nuances of what it means to say that a property is typical. It should be understood as a technical term, capturing that these laws are derived from a typical-

[7] The idea that typicality predicts membership could be captured by adding an additional constraint to the model: if $x \succeq_\Phi y$ and $y \in I(\Phi)$, then $x \in I(\Phi)$, implying that non-members are always less typical than members.

ity ordering, which itself is based on a prototype. They can be understood as laws of normality (Φ are normally Ψ) or non-exceptionality (Non-exceptional Φ are Ψ).

The relation of $\mathcal{T}\Phi\Psi$ to $\mathcal{D}\Phi\Psi$ is obvious. The truth conditions of $\mathcal{T}$ weaken those of $\mathcal{D}$ with the result that $\mathcal{T}\Phi\Psi$ is a logical consequence of $\mathcal{D}\Phi\Psi$. What holds for all plausible members holds also for the plausible members that are typical.

The following valid inferences, which immediately follow from set-theoretic notions, are worth mentioning:

1. $\mathcal{T}\Phi\Psi$ implies $\mathcal{T}\Phi\Psi \cup \Xi$.
   For every $\mathbf{M}$ such that $T_{\mathbf{M}}(\Phi) \subseteq I(\Psi)$: $T_{\mathbf{M}}(\Phi) \subseteq I(\Psi \cup \Xi) = I(\Psi) \cup I(\Phi)$.
2. $\mathcal{T}\Phi\Psi$ and $\mathcal{T}\Phi\Xi$ implies $\mathcal{T}\Phi\Psi \cap \Xi$
   For every $\mathbf{M}$ such that $T_{\mathbf{M}}(\Phi) \subseteq I(\Psi)$ and $T_{\mathbf{M}}(\Phi) \subseteq I(\Xi)$: $T_{\mathbf{M}}(\Phi) \subseteq I(\Psi \cap \Xi) = I(\Psi) \cap I(\Xi)$.

The first inference – a version of the so-called weakening of consequent – says that if a property is true for typical members, then any extension of the property is typically true as well. If typical birds have a beak, then they have a beak *or* teeth. Whether the argument scheme is intuitively plausible can be debated.

Veltman [31] endorses that his system of normality laws does *not* validate weakening of the consequent. This is due to the dynamic character of his normality laws, where the weakened statement leads to a different update in the expectations: "'Tigers normally have four or five legs' indicates what one can expect in case one encounters a tiger that does not have four legs; the rule 'Tigers normally have four legs' does not" [31, 257]. Our typicality laws, however, do not *induce* expectations but only *express* them.

Another more general criticism one can raise against weakening of consequent lies in the word "typical". Shouldn't typical properties not only be expected but also to some degree diagnostic, i.e., indicate membership? However, expectedness and diagnosticity have inverse logical properties. In terms of probability theory, high $P(\Psi|\Phi)$ measures the expectedness of the property Ψ and high $P(\Phi|\Psi)$ its diagnosticity for category Φ.[8] Our typicality laws are not at all about diagnosticity. However, the above given semantics for $\mathcal{T}$ could be used to define more restricted kinds of typicality laws which incorporate some intuitions about the diagnosticity of typicality. One could, for example, require that a typical property should not be typical in the complement, i.e., non-trivial: $\mathcal{T}\Phi\Psi \wedge \neg\mathcal{T} - \Phi\Psi$. It is also possible to demand that typical properties are atypical in the complement: $\mathcal{T}\Phi\Psi \wedge \mathcal{T} - \Phi - \Psi$. Finally, one can define two directional typicality laws: $\mathcal{T}\Phi\Psi \wedge \mathcal{T}\Psi\Phi$. Within this paper, we will not go into the details of these notions of typicality, which do not validate weakening of consequent. $\mathcal{T}$, as defined above, suffices to express different variations of typicality laws.[9]

After emphasizing that our typicality notion is about what one expects, it is important to note that it is not a probabilistic one. This becomes clear in the second

[8] Schurz [28] calls the former "typicality in the wider sense" and the combination of the former and second "typicality in the narrow sense".

[9] In particular, $\mathcal{T}$ is sufficient to formulate different preconditions of rational predicate change.

inference, a set-theoretic version of the AND rule. It says that one can agglomerate typical properties by intersection and that the result will still be a typical property. The AND rule is a probabilistically lossy inference. If at least 99 % of mammals are viviparous and at least 99% of mammals are quadrupeds, it is not guaranteed that at least 99 % of them are both. There is a lower limit though: the extension of mammals consists of at least 98% viviparous quadrupeds.[10] In a nutshell, the ADT account does not identify typicality with relative frequencies but it is compatible with accounts that take statistical frequency as an important component of typicality [27, 29].

4 Predicate Change in AD and ADT

We will now discuss how the change of atomic predicates, i.e., a change of analytic laws, can be captured and expressed by an enrichment of the outlined logics AD and ADT. Since category change involves alteration and continuity, there are two versions of predicate change: new members are included but no member vanishes or members are excluded but no new member enters the category. Moreover, predicate change involves not only a predicate that is changed but also a criterion that determines its result. From these remarks, we suggest the following definitions of exclusive and inclusive predicate change:

Definition 11 (Inclusion) Let Ψ be an atomic predicate. The inclusion of Ψ with respect to Φ on $\mathbf{M}$ yields the model $\mathbf{M}|\Psi \uparrow \Phi$, where $I \in \mathbf{M}$ is changed to $I' \in \mathbf{M}|\Psi \uparrow \Phi$ with $I'(\Psi) = I(\Psi \cup \Phi)$ and $I'(\Xi) = I(\Xi)$ for all other atomic predicates Ξ.

Definition 12 (Exclusion) Let Ψ be an atomic predicate. The exclusion of Ψ with respect to Φ on $\mathbf{M}$ yields the model $\mathbf{M}|\Psi \downarrow \Phi$, where $I \in \mathbf{M}$ is changed to $I' \in \mathbf{M}|\Psi \downarrow \Phi$ with $I'(\Psi) = I(\Psi \cap \Phi)$ and $I'(\Xi) = I(\Xi)$ for all other atomic predicates Ξ.

These two definitions provide the foundation for extending AD and ADT to a dynamic logic of analytic change. We introduce two operators that express the results of the semantic change:

Definition 13 (Dynamic AD and ADT) Dynamic AD and dynamic ADT add two unary modal operators to AD and ADT, respectively:

1. $[\Psi \uparrow \Phi]\phi$ and $[\Psi \downarrow \Phi]\phi$ are sentences of dynamic AD / ADT iff
   (a) Ψ is an atomic predicate of AD /ADT,
   (b) Φ is a (possibly complex) predicate of AD /ADT, and
   (c) ϕ is a sentence of dynamic AD / ADT.

[10] For more details on lossy inferences, see [22].

2. $\mathbf{M} \models [\Psi \uparrow \Phi]\phi$ iff $\mathbf{M}|\Psi \uparrow \Phi \models \phi$, and
   $\mathbf{M} \models [\Psi \downarrow \Phi]\phi$ iff $\mathbf{M}|\Psi \downarrow \Phi \models \phi$

Note that the dynamic operators can be iterated and occur inside sentences. For example, $[\Psi \downarrow \Phi]\phi \wedge [\Phi \uparrow (\Xi_1 \cap \Xi_2)][\Phi \downarrow \Xi_3]\psi$ is a scheme of a sentence, where Ψ and Φ need to be atomic predicates.

The first thing to address is the relation between AD / ADT and their dynamic counterparts. In the case of AD, this is quite straightforward: dynamic AD is reducible to AD.

Proposition 1 (Reduction of dynamic AD to AD) *Every formula of dynamic AD can be translated to a formula of AD by the following two equivalences, formulated for an atomic Ξ_1 that is changed with respect to Ξ_2:*

I *If ϕ contains no operation of predicate change, then $[\Xi_1 \uparrow \Xi_2]\phi$ is equivalent to $\phi_{\lceil \Xi_1 \cup \Xi_2/\Xi_1 \rceil}$ the result of substituting every Ξ_1 by $\Xi_1 \cup \Xi_2$ in ϕ.*

E *If ϕ contains no operation of predicate change, then $[\Xi_1 \downarrow \Xi_2]\phi$ is equivalent to $\phi_{\lceil \Xi_1 \cap \Xi_2/\Xi_1 \rceil}$ the result of substituting every Ξ_1 by $\Xi_1 \cap \Xi_2$ in ϕ.*

In case several action expressions are nested, one starts at the most inner one and eliminates them step by step.

Proof We prove this by induction. In the base step, we show that the equivalence holds for the atomic formulae, $\mathcal{A}\Phi\Psi$ and $\mathcal{D}\Phi\Psi$. The induction step shows that the equivalence is then inherited to formulae that are more complex.

– For $\mathcal{A}\Phi\Psi$: $[\Xi_1 \uparrow \Xi_2]\mathcal{A}\Phi\Psi$ iff $\mathcal{A}\Phi_{\lceil \Xi_1 \cup \Xi_2/\Xi_1 \rceil}\Psi_{\lceil \Xi_1 \cup \Xi_2/\Xi_1 \rceil}$. By $[\Xi_1 \uparrow \Xi_2]$ one refers to a model $\mathbf{M}|A$ in which I is changed to I' such that $I'(\Xi_1) = I(\Xi_1 \cup \Xi_2)$. The definition of $\mathcal{A}$ only depends on I and I', which yield the same sets before and after the change, respectively.
  $[\Xi_1 \downarrow \Xi_2]\mathcal{A}\Phi\Psi$ iff $\mathcal{A}\Phi_{\lceil \Xi_1 \cap \Xi_2/\Xi_1 \rceil}\Psi_{\lceil \Xi_1 \cap \Xi_2/\Xi_1 \rceil}$ follows similarly.
– For $\mathcal{D}\Phi\Psi$:
  $[\Xi_1 \uparrow \Xi_2]\mathcal{D}\Phi\Psi$ iff $\mathcal{D}\Phi_{\lceil \Xi_1 \cup \Xi_2/\Xi_1 \rceil}\Psi_{\lceil \Xi_1 \cup \Xi_2/\Xi_1 \rceil}$. I is changed to I' as above. $\mathcal{D}$ depends not only on I and I' but also on the plausibility relation $\geqslant$. However, $P_{\mathbf{M}|\mathbf{A}}(\Xi_1)$ is still identical to $P_{\mathbf{M}}(\Xi_1 \cup \Xi_2)$ because the plausibility relation is in both models identical. The changes in the set of plausible members thus only depend on the change of I to I'.
  $[\Xi_1 \downarrow \Xi_2]\mathcal{D}\Phi\Psi$ iff $\mathcal{D}\Phi_{\lceil \Xi_1 \cap \Xi_2/\Xi_1 \rceil}\Psi_{\lceil \Xi_1 \cap \Xi_2/\Xi_1 \rceil}$ follows similarly.
– Since the connectives are extensional, it holds: if ϕ_1 is equivalent to ϕ_2 and ψ_1 is equivalent to ψ_2, then $\neg\phi_1$ is equivalent to $\neg\phi_2$, $\phi_1 \wedge \psi_1$ is equivalent to $\phi_2 \wedge \psi_2$, and $\phi_1 \vee \psi_1$ is equivalent to $\phi_2 \vee \psi_2$.
– Every formula ϕ without predicate change operations only consists of $\mathcal{D}\Phi\Psi$, $\mathcal{A}\Phi\Psi$, and the propositional connectives. Thus, $[\Phi \downarrow \Psi]\phi$ is equivalent to $\phi_{\lceil \Phi \cap \Psi/\Phi \rceil}$ and $[\Phi \uparrow \Psi]\phi$ is equivalent to $\phi_{\lceil \Phi \cup \Psi/\Phi \rceil}$ if ϕ contains no predicate changing operation.

– If ϕ contains predicate changes, the subformulae without further nested predicate changes are translated before the next outer operations until all predicate changes are eliminated. □

The possibility of translation shows that predicate change, while being a kind of conceptual change, has no effects on expressive power and is far from causing incommensurability. The analytic and doxastic laws that become true for Φ were previously true for $\Phi \cup \Psi$ or $\Phi \cap \Psi$ respectively. Considering Carnap's example of the changed predicate FISH, this means that after excluding whales from fish, the new laws about fish are the old laws about fish that are not whales.

A corresponding proposition is not valid for ADT, because typicality laws are concept dependent. After excluding whales from fish, the new typical properties of fish are not the previous typical properties of fish that are not whales, because we do not have *proper* typicality laws for FISH THAT ARE NO WHALES.[11] Thus, $\mathcal{T}$ laws are not translatable from dynamic ADT to ADT.

There is a way to overcome this untranslatability. With a minor enrichment of ADT one can get an according reduction. We need to introduce ternary laws $\mathcal{T}^{\Xi}\Phi\Psi$, read as Ψ is Ξ-typical for Φ, extending (dynamic) ADT to (dynamic) ADT3.

Definition 14 (ADT3 and ternary $\mathcal{T}$ laws) The logic ADT3 is defined as an extension of ADT that includes ternary typicality laws:

All predicates and formulae of ADT are predicates and formulae of ADT3.
If Ξ, Φ, and Ψ are predicates of ADT3 , then $\mathcal{T}^{\Xi}\Phi\Psi$ is a formula of ADT3.
$\mathbf{M} \models \mathcal{T}^{\Xi}\Phi\Psi$ iff $T^{\Xi}_{\mathbf{M}}(\Phi) \subseteq I(\Psi)$, where $T^{\Xi}_{\mathbf{M}}(\Phi) = \{x \in P_{\mathbf{M}}(\Phi) | \neg\exists y (y \in P_{\mathbf{M}}(\Phi) \wedge y \succ_{Pt(\Xi)} x)\}$.
Dynamic ADT3 is the extension of ADT3 by the dynamic operations as outlined in Defintion 13.

Put in words, $\mathcal{T}^{\Xi}\Phi\Psi$ is true if and only if Ψ is true for those plausible Φ that resemble the prototype of a third predicate Ξ best. With this law type, one can, for example, express that the most fish-like mammals (cetaceans) live in the ocean or that the most reptile-like mammals (monotremes, e.g., platypus) lay eggs. The usual typicality laws, we introduced previously, are just a special case where $\Xi = \Phi$.

The most important thing about dynamic ADT3 is that it allows for translation between typicality laws before and after a conceptual change:

Proposition 2 *Every formula $[\Xi_1 \uparrow \Xi_2]\phi$ and $[\Xi_1 \downarrow \Xi_2]\phi$ in dynamic ADT3 can be translated to an equivalent formula without action operators by Proposition 1 and the following two equivalences:*

– *$[\Xi_1 \uparrow \Xi_2]\mathcal{T}^{\Xi}\Phi\Psi$ is equivalent to $\mathcal{T}^{\Xi}\Phi_{[\Xi_1 \cup \Xi_2/\Xi_1]}\Psi_{[\Xi_1 \cup \Xi_2/\Xi_1]}$, the result of substituting any Ξ_1 by $\Xi_1 \cup \Xi_2$ in Φ and Ψ.*

[11] Since the typicality ordering cannot discriminate between different members, all plausible members are typical. The only typicality laws follow from corresponding doxastic laws.

– *$[\Xi_1 \downarrow \Xi_2]\mathcal{T}^{\Xi}\Phi\Psi$ is equivalent to $\mathcal{T}^{\Xi}\Phi_{\ulcorner\Xi_1\cap\Xi_2/\Xi_1\urcorner}\Psi_{\Xi_1\cap\Xi_2/\Xi_1}$, the result of substituting any Ξ_1 by $\Xi_1 \cap \Xi_2$ in Φ and Ψ.*

Proof By Proposition 1, we can translate laws of the form $\mathcal{A}\Phi\Psi$ and $\mathcal{D}\Phi\Psi$. Binary typicality laws $\mathcal{T}\Phi\Psi$ can be reformulated to $\mathcal{T}^{\Phi}\Phi\Psi$. We show that this ternary law is indeed translatable by proving the above-given equivalences:

The truth value of $\mathcal{T}^{\Phi}\Phi\Psi$ depends on $I(\Psi)$ and on $T^{\Phi}_{\mathbf{M}}(\Phi)$. In order to determine the set $T^{\Phi}_{\mathbf{M}}(\Phi)$, one first needs to fix the plausible members of Φ. This works as outlined for $\mathcal{D}$ above. Now, we only need to ensure that the typicality ordering remains the same. This is guaranteed by the fact that reference of the typicality, i. e. Ξ, is fixed. The translation of $I(\Psi)$ is trivial.

The equivalences for ternary $\mathcal{T}$ laws, together with the ones for $\mathcal{A}$ and $\mathcal{D}$ from Proposition 1, suffice to translate all basic laws. For connectives and nesting the proof proceeds as above for AD and ADT. □

Predicate change in dynamic ADT is not incommensurable in the sense that there can be no communication about laws before and after the change. All formulae can be translated to dynamic ADT3, where dynamic formulae have an equivalent non-dynamic formula. Note, however, that binary typicality laws have only corresponding ternary translations. For example, after excluding WHALES from FISH, the new typicality laws for FISH are not the old typicality laws for FISH THAT ARE NOT WHALES, but the laws of fish-likeness for FISH THAT ARE NO WHALES. The formulation of ternary typicality is slightly weird and uncommon in natural language. In this sense, typicality laws are more difficult to translate. Luckily, as we discuss in the next section, many predicate changes do not affect typical members and we will not have to rely on translations to ternary typicality laws.[12]

Predicate change entails no change in expressive power. However, it influences the laws of the system, most directly the analytic laws.[13] It is easily verified that the following two formulae are tautological:

$$[\Phi \uparrow \Psi]\mathcal{A}\Psi\Phi$$
$$[\Phi \downarrow \Psi]\mathcal{A}\Phi\Psi$$

It might be worrying that this holds even for seemingly inconsistent analytic laws. For example, $\Phi \uparrow -\Phi$ empties $I'(-\Phi)$, which makes $\mathcal{A} - \Phi\Phi$ trivially true. A similar effect occurs for $\Phi \downarrow \Phi$: now $I'(\Phi) = \emptyset$, which trivially fulfils $\mathcal{A}\Phi - \Phi$. However, such trivialisations only occur if a meaningful predicate is changed to a paradoxical one (i.e., a predicate without conceivable instances).

Any analytic law, even for complex predicates, can be created by procedures of inclusion and exclusion, as demonstrated by Proposition 7 in the Appendix. The general applicability of predicate change is a virtue in its own right, but it does not

[12] We will no longer consider ADT3 and ternary laws, which we only introduced in order to secure translatability, in the next section.

[13] From here on, we don't compare equivalent translations, but the laws before and after the change, e.g., $\mathcal{A}\Phi\Psi$ with $[\Xi_1 \uparrow \Xi_2]\mathcal{A}\Phi\Psi$ and $[\Xi_1 \downarrow \Xi_2]\mathcal{A}\Phi\Psi$

capture the rational restrictions predicate change should have and usually has. This will be the central theme of the following section.

5 Rational Restrictions

There are two ways to evaluate the rationality of predicate change. We can look at the initial state of the belief system that encourages predicate change, i.e., its motivations. The second option is to evaluate the results of predicate change, especially whether the new system preserves laws of the old system. We begin by sketching the first perspective before giving attention to the second one.

5.1 Motivation

An important motivation of predicate change is an adjustment of meaning to doxastic laws or to typicality laws by excluding category members that violate the law. We will call such a change *analytification by exclusion*. Its doxastic version presupposes a law $\mathcal{D}\Phi\Psi$, while the typicality driven exclusion is motivated by $\mathcal{T}\Phi\Psi$.

Definition 15 (Analytification by exclusion) The exclusion $\Phi \downarrow \Psi$ on model $\mathbf{M}$ is a doxastic analytification iff $\mathbf{M} \models \mathcal{D}\Phi\Psi$ and a typicality analytification iff $\mathbf{M} \models \mathcal{T}\Phi\Psi$.

The doxastic analytification is appropriate if a law is quite firm and an empirical revision unwarranted. For example, since we are highly confident that humans are mortal, we might refuse to call an immortal human still a human, even if begotten by humans. By excluding this, we create a conceptual law from a doxastic one. Doxastic analytification only concerns hypothetical category members. It adjusts conceptual systems to fundamental beliefs, but it neither covers the recent exclusion of Pluto from the category of planets nor the reclassification of whales, as depicted by Carnap. In both cases, real, maximally plausible members were excluded. Their "defect" was rather that they differed from the more paradigmatic instances in crucial aspects. They were *atypical* in some of their properties.

Typicality analytification by exclusion makes some typical properties analytical. Within this process, several atypical members are excluded from the category. The reasons for this step can be diverse. In the case of Pluto, the new definition – a round celestial object that orbits a star and has cleared its neighbourhood – was arguably motivated by the wish to keep the number of existing planets low. The redefinition seemed slightly artificial and raised criticism. In Carnap's example, however, the new defining properties of fish, like inner body structure or gills, became indeed more central.

Analytification by inclusion looks different. If we took $\mathcal{T}\Phi\Psi$ or $\mathcal{D}\Phi\Psi$ as justification for $\Phi \uparrow \Psi$, we would create a reverse analytic law of the form $\mathcal{A}\Psi\Phi$. This turns a plausible or typical property into a sufficient condition of membership, which is highly questionable. For example, humans are mortal but it would be very strange to call every mortal being a human. For this kind of change, we rather need a law $\mathcal{D}\Psi\Phi$ or $\mathcal{T}\Psi\Phi$, where the changed predicate occurs as the ascribed property. Thus, one may extend the meaning of a predicate Φ to $\Phi \cup \Psi$ if the most plausible or the most typical Ψ are already Φ.

Definition 16 (Analytification by inclusion) The inclusion $\Phi \uparrow \Psi$ on model $\mathbf{M}$ is a doxastic analytification iff $\mathbf{M} \models \mathcal{D}\Psi\Phi$ and a typicality analytification iff $\mathbf{M} \models \mathcal{T}\Psi\Phi$.

Analytifications are not the only kinds of predicate change that can occur. We will look at a further kind of change, namely the integration of categories into superordinated categories by inclusive predicate change. For example, mammals plausibly have a number of crucial features (middle ear bones, diaphragma and so on). A category the members of which expose these crucial features as well is perceived as belonging to mammals. The reasoning resembles an abduction. Subsuming the subcategory under another category explains many of its crucial features. Here the motivation behind $\Phi \uparrow \Psi$ is that fundamental laws of Φ apply to Ψ as well and cover several of its regularities. If $\mathcal{D}\Phi\Xi_1$, $\mathcal{D}\Phi\Xi_2$,..., $\mathcal{D}\Phi\Xi_n$ are the crucial laws of Φ and it also holds that $\mathcal{D}\Psi\Xi_1$, $\mathcal{D}\Psi\Xi_2$,..., $\mathcal{D}\Psi\Xi_n$, then it is appropriate to perceive Ψ as a subcategory of Φ.[14] Let us thus say that integration is relative to a critical predicate Ξ, usually an intersection of important properties (i.e.,$\Xi = \Xi_1 \cap \Xi_2 \cap \ldots \cap \Xi_n$), that serves as a criterion of the integration.

The same considerations can be made for typicality laws. If central typical properties of Φ are also typical for Ψ, we can view Ψ as subcategory of Φ. In folk taxonomies, typicality integration can even overrule scientific classification. For example, instead of their close biological relatedness, melons are integrated into the fruit category while cucumbers are considered as legume, because their typical tastes and usages resemble these different categories, respectively.

We define the basic integration for doxastic and typicality laws accordingly:

Definition 17 (Integration) For an atomic predicate Φ the inclusion $\Phi \uparrow \Psi$ on model $\mathbf{M}$ is a doxastic integration of Ψ to Φ with respect to Ξ if $\mathbf{M} \models \mathcal{D}\Phi\Xi$ as well as $\mathbf{M} \models \mathcal{D}\Psi\Xi$ and a typicality integration with respect to Ξ if $\mathbf{M} \models \mathcal{T}\Phi\Xi$ as well as $\mathbf{M} \models \mathcal{T}\Psi\Xi$.

This definition gives the fundamental procedure of integration. However, our intuition about justified integrations is more demanding, especially with respect to the criterion Ξ. For one thing, the role of the criterion is different for Φ and Ψ. The criterion should be very salient for Φ while Ψ can have many other, even more important properties. Intuitively, integration should not be symmetric. An integration of MELON

[14]Note, however, that we should not demand that this relationship holds for *all* $\mathcal{D}$ laws of Φ. This would also presuppose that $\mathcal{D}\Psi\Phi$, since $\mathcal{D}\Phi\Phi$ is logically true, and integration would collapse into analytification.

to FRUIT means that melons become a subcategory of fruits. This is something totally different from the (slightly insane) integration of FRUIT into MELON.

There are several constraints one can impose on Ξ, the criterion of integration. First, one can demand that Ξ comprises all non-analytic regularities of Φ. This intuitive condition can be called *centrality*. Another way to emphasize the specific role Ξ plays for Φ is to demand that Φ is the only predicate that is lawfully related to Ξ. No conceptually disjoint predicate accords to the same laws. This demand can be called *distinctiveness*. Finally, if Ξ is the criterion of integrating Ψ, it should not lawfully apply to other predicates that are *not* integrated. We call this demand *exhaustiveness*. These additional restrictions quantify over predicates. They are expressible as metalogical conditions. We can formulate them for doxastic integrations as well as for typicality integrations

- Centrality
  - A doxastic integration $\Phi \uparrow \Psi$ on $\mathbf{M}$ with respect to Ξ is central iff for all predicates Υ such that $\mathcal{D}\Phi\Upsilon \wedge \neg\mathcal{A}\Phi\Upsilon$ it holds that $\mathcal{D}\Phi\Xi \models \mathcal{D}\Phi\Upsilon$.
  - A typicality integration $\Phi \uparrow \Psi$ on $\mathbf{M}$ with respect to Ξ is central iff for all predicates Υ such that $\mathcal{T}\Phi\Upsilon \wedge \neg\mathcal{A}\Phi\Upsilon$ it holds that $\mathcal{T}\Phi\Xi \models \mathcal{T}\Phi\Upsilon$.[15]
- Distinctiveness
  - A doxastic integration $\Phi \uparrow \Psi$ on $\mathbf{M}$ with respect to Ξ is distinctive iff for all predicates Υ, such that $\mathbf{M} \models \mathcal{A}\Phi - \Upsilon$, it holds that $\mathbf{M} \models \neg\mathcal{D}\Upsilon\Xi$.
  - A typicality integration $\Phi \uparrow \Psi$ on $\mathbf{M}$ with respect to Ξ is distinctive iff for all predicates Υ, such that $\mathbf{M} \models \mathcal{A}\Phi - \Upsilon$, it holds that $\mathbf{M} \models \neg\mathcal{T}\Upsilon\Xi$.
- Exhaustiveness
  - A doxastic integration $\Phi \uparrow \Psi$ on $\mathbf{M}$ with respect to Ξ is exhaustive iff for all predicates Υ, such that $\mathbf{M} \models \mathcal{D}\Upsilon\Xi$, it holds that $\mathbf{M} \models \mathcal{A}\Upsilon\Psi$.
  - A typicality integration $\Phi \uparrow \Psi$ on $\mathbf{M}$ with respect to Ξ is exhaustive iff for all predicates Υ, such that $\mathbf{M} \models \mathcal{T}\Upsilon\Xi$, it holds that $\mathbf{M} \models \mathcal{A}\Upsilon\Psi$.

Reconsidering Carnap's example of the whales, the critical change is not so much the exclusion of whales from fish but their integration to mammals, based on the fact that properties for mammals – the suckling of the offspring, the structure of the heart, lungs, etc. – were known to be true about whales as well.

Let us characterise this integration in terms of the given constraints. The inclusion of whales to mammals in the tenth edition of *Systema Naturae* is distinctive. It is also exhaustive insofar as Linnaeus included all (known) species with the respective properties to the category of mammals. No other disjoint category had the same

[15] With this formulation, the typicality integration also includes a doxastic integration, because all doxastic laws are also typicality laws. One can define a weaker version of typicality integration by presupposing $\mathcal{T}\Phi\Upsilon \wedge \neg\mathcal{D}\Phi\Upsilon$ instead of $\mathcal{T}\Phi\Upsilon \wedge \neg\mathcal{A}\Phi\Upsilon$.

properties. Arguably, the integration was central or at least close to central, as it was based on the proper plausible properties of mammals that were known. The seeming counterexample, namely that mammals are quadrupeds, was already loosened up when Linnaeus decided to include humans to the quadrupeds in the first edition of *Systema Naturae*.[16]

By classifying cetaceans as mammals, one also needed to exclude them from the fish. Otherwise, the analytic truth that fish are no mammals would be lost. In this case, one predicate change motivated another one. As a final motivation of predicate change we define such a recovering predicate change:

Definition 18 (Recovering) For an atomic Φ and a model $\mathbf{M}|A$ after a change A, the inclusion $\Phi \uparrow \Psi$ on $\mathbf{M}|A$ is a recovery of $\mathcal{A}\Psi\Phi$ iff $\mathbf{M} \models \mathcal{A}\Psi\Phi$, and the exclusion $\Phi \downarrow \Psi$ on $\mathbf{M}|A$ is a recovery of $\mathcal{A}\Phi\Psi$ iff $\mathbf{M} \models \mathcal{A}\Phi\Psi$.

For the predicate change of WHALE, we propose the following reconstruction, which slightly differs from the one Carnap gives. After having been perceived as a kind of fish, whales were integrated into the category of mammals on the ground of numerous unique features they share with mammals but not with other animals. Linaeaus names inter alia the structure of the heart, the warm blood, the lungs, jaws, lactiferous teats in females, and hair. The exclusion of cetaceans from fish was merely a recovering predicate change, directly following the previous integration. Within the whole process, the outer appearance, as well as the habitat of the animal, lost importance. An aquatic living environment and fish-like shape was no longer a sufficient condition for *being* a fish. The formerly analytic property of quadrupedia to have four feet was lost in the category of mammals, but it remained a typicality law. Linné [20] mentions having four legs and a tail as typical properties of mammals. Thus, even within the quite drastic restructuring of taxonomy, we find continuity. This brings us to the second perspective we can take when evaluating conceptual change, namely conservativity. To which degrees do exclusions and inclusions preserve doxastic and typicality laws?

5.2 Conservativity

Let us now discuss restrictions of a predicate change A, either inclusion or exclusion, by its impact on the non-analytic laws. That means, we compare the status of the laws $\mathcal{D}\Phi\Psi$ and $\mathcal{T}\Phi\Psi$ in $\mathbf{M}$ and $\mathbf{M}|A$. A completely radical change can lead to an alteration of plausible and typical members of Φ by new ones. It would be possible to replace laws like $\mathcal{T}\Phi\Psi$ and $\mathcal{D}\Phi\Psi$ by the contrary laws $\mathcal{T}\Phi - \Psi$ and $\mathcal{D}\Phi - \Psi$. The idea of conservativity is that the impact of conceptual change is restricted.

A minimally restricted form of predicate change is a *slightly conservative* predicate change. This excludes that laws are replaced by contrary ones but allows that old laws are lost *and* that new laws begin to hold. By a so-called *moderately conserva-*

[16] Digital versions of the editions are available in the biodiversity heritage library: [21] first included humans as quadrupeds and [20] included the cetaceans and renamed the category to “mammalia”.

tive change, we demand that only one of this happens. Either new laws arise but old ones persist, or old laws are lost but no new laws arise. The maximally conservative version is *strictly conservative* predicate change, which forbids both, the loss of old laws and the arising of new laws for the changed predicate.

Conservativity restrictions can also be formulated by the effect a predicate change has on the plausible and typical members of the category in model **M** in comparison to $\mathbf{M}|A$. Let us characterise this for typicality. Non-conservative change allows that all typical members are replaced, either because they are excluded or because new members suppress the old ones. A slightly conservative change forbids this. For a moderate change, we demand that the old typical members are either a subset or a superset of the new members. Strictly conservative change requires that there is no change of typical members at all. Table 1 gives an overview of increasingly demanding degrees of conservativity for typicality laws and typical members.

The according characterisation can be applied for plausible members. There is, however, one notable deviation. Because the plausibility ordering is totally connected, the slightly conservative change of plausible members collapses into the moderate variant. If one of the maximal members is degraded by an inclusive predicate change, then all others will be degraded as well. Similarly for exclusion: if a formerly dominated member becomes upgraded after exclusion of some maximal members, all maximal members must have been excluded. If one maximal member remains after the change, no previously dominated object can rise to this status. In both cases, any slightly conservative change is also moderately conservative. Table 2 gives the corresponding overview.

The conservativity constraints for laws and for members are not equivalent, though they are formulated in a corresponding way. The restriction on the typical or plausible members implies the according restriction on the laws. For example, if the typical members are the same before and after the change, then the typicality laws will be the same before and after the change because they are defined on the set of typical members. However, it is possible that there is no change in typicality laws but a change in the typical members, namely if the new typical members comply with the same laws as the old ones.

The preservation of laws by conservative predicate change on a predicate Φ can be illustrated by the relations of the square of oppositions. Slightly conservative change

Table 1 Conservativity restrictions for Φ-changing action A (expressed by $[A]$) on **M** with respect to the *typicality laws* of Φ and with respect to the *typical members* of Φ

Restriction	For typicality laws	For typical members
Slight conservativity	If $\mathcal{T}\Phi\Psi$, then $[A]\neg\mathcal{T}\Phi - \Psi$.	$T_{\mathbf{M}}(\Phi) \cap T_{\mathbf{M}\|A}(\Phi) \neq \emptyset$
Moderate conservativity...		
...for Inclusion	If $\neg\mathcal{T}\Phi\Psi$, then $[\Phi \uparrow \Psi]\neg\mathcal{T}\Phi\Psi$.	$T_{\mathbf{M}}(\Phi) \subseteq T_{\mathbf{M}\|A}(\Phi)$.
...for Exclusion	If $\mathcal{T}\Phi\Psi$, then $[\Phi \downarrow \Psi]\mathcal{T}\Phi\Psi$.	$T_{\mathbf{M}}(\Phi) \supseteq T_{\mathbf{M}\|A}(\Phi)$.
Strict conservativity	$\mathcal{T}\Phi\Psi$ iff $[A]\mathcal{T}\Phi\Psi$.	$T_{\mathbf{M}}(\Phi) = T_{\mathbf{M}\|A}(\Phi)$.

Table 2 Conservativity restrictions for Φ changing action A (expressed by $[A]$) on $\mathbf{M}$ with respect to the *doxastic laws* of Φ and with respect to the *plausible members* of Φ

Restriction	For doxastic laws	For plausible members
Slight conservativity	If $\mathcal{D}\Phi\Psi$, then $[A]\neg\mathcal{D}\Phi-\Psi$.	
Moderate conservativity...		
...for Inclusion	If $\neg\mathcal{D}\Phi\Psi$, then $[\Phi\uparrow\Psi]\neg\mathcal{D}\Phi\Psi$.	$P_{\mathbf{M}}(\Phi)\subseteq P_{\mathbf{M}\vert A}(\Phi)$.
...for Exclusion	If $\mathcal{D}\Phi\Psi$, then $[\Phi\downarrow\Psi]\mathcal{D}\Phi\Psi$.	$P_{\mathbf{M}}(\Phi)\supseteq P_{\mathbf{M}\vert A}(\Phi)$.
Strict conservativity	$\mathcal{D}\Phi\Psi$ if and only if $[A]\mathcal{D}\Phi\Psi$.	$P_{\mathbf{M}}(\Phi)=P_{\mathbf{M}\vert A}(\Phi)$.

The cell from slightly conservative change of members is empty, since such a change of plausible members is always moderatly conservative as well

allows going from a universal operation to its particular dual one. Moderate exclusion preserves (universal) laws and moderate inclusion their (particular) duals. Strict change preserves the laws in all corners of the square. Figure 1 shows a prior and a posterior square of opposition for doxastic laws. The strictly conservative preservation is marked with grey arrows. Moderate conservativity is given in dashed arrows with the finer dashing for exclusion. The dotted arrows represent the preservation by slightly conservative change. An according representation can be applied for typicality laws.

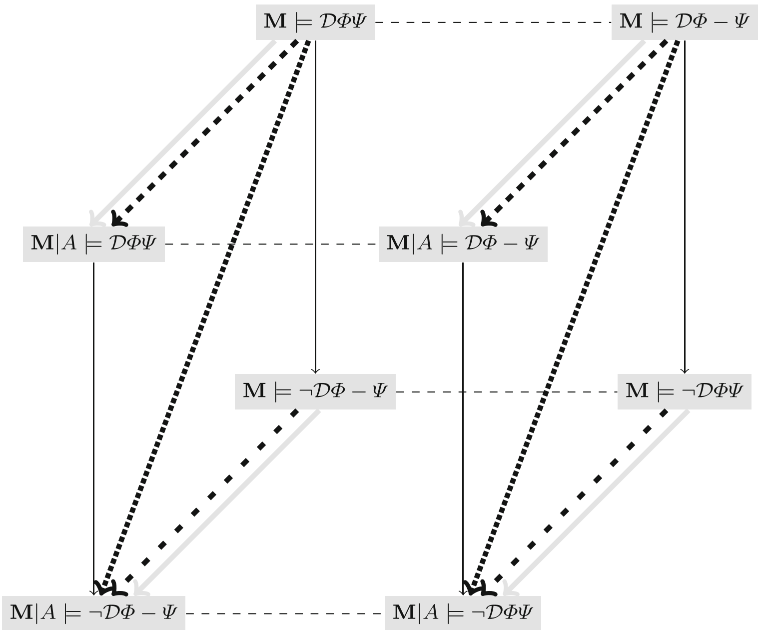

Fig. 1 Preservation in a Square of Opposition: The old square is in the background. Grey arrows represent strictly conservative change. Dashed arrows represent moderate predicate change. The dotted arrow represents slightly conservative predicate change

It is noteworthy that analytications, as defined above, are strictly conservative. The typicality analytification can only affect non-typical members of the category and thus preserves its typicality laws. Doxastic laws, however, can be affected. For example, the typicality exclusion of the now so-called dwarf planets from the category of planets had an impact on Pluto, a real and thus plausible entity, which fell under the former definition of planet. A doxastic analytification, on the other hand, only affects non-plausible members and is thus strictly conservative with respect to doxastic laws *and* with respect to typicality laws.

For an integration $\Phi \uparrow \Psi$ with respect to Ξ, there is at least a tendency towards strict conservativity. It is clear that $\mathcal{D}\Phi\Xi$ (in case of a doxastic integration) or $\mathcal{T}\Phi\Xi$ (in case of a typicality integration), as well as everything that follows from that law, is preserved. If many plausible properties of Φ are included in Ξ, many doxastic laws will be preserved. Thus, while approaching the demand of centrality, integrations also approach strict conservativity.

Strictly conservative predicate change is philosophically interesting. It affects *only* the conceptual beliefs about the category members. If you and I start with the same beliefs and concepts but change the same category strictly conservatively in different directions, we still share beliefs about the predicate, though we disagree on its meaning. For example, the classification of Pegasus as horse or non-horse is irrelevant for our beliefs about real horses. Disagreement in analytic laws is often merely disagreement on objects we do not believe to exist. A strictly conservative category change can be thus almost unrecognisable.

6 Conclusion

This paper discussed three different kinds of laws. Analytic laws are universal truths for all possible instances of a concept, including absurd but imaginable members of the category. Second, laws that are universally true for plausible instances of a concept were called "doxastic laws". Exceptions are imaginable but implausible. Finally, we introduced typicality laws that allow for real exceptions, for example, "Birds can fly". We modeled how analytic laws are revised by predicate changing actions. While these operations can be used to create analytic laws in a more or less arbitrary way, we also argued that predicate change is usually justified by pre-existing laws and restricted by a need for continuity. The formal results of the paper contribute to two philosophical issues: the severity of conceptual change and the status of analytic laws.

Often conceptual change is associated with scientific revolutions [30] or major steps in the cognitive development of young children [7]. Since [16], such scientific revolutions were linked to incommensurability, i.e., a lack of common language.[17] Our system deals with a milder form of conceptual change. Predicate change causes no alteration of expressive possibilities. The laws of the old system have translations

[17] The notion of incommensurability is of course far more complex and there is no commonly agreed way to understand it, but all of the interpretations assume a lack of translatability from one system to the other system.

in the new system. Every doxastic statement after a predicate change is directly translatable to an equivalent statement before the change. For typicality laws, however, this can only be achieved by a ternary typicality law.

Rationally justified predicate change is restricted by conservativity. In its most strict form, conservativity entails that all typicality laws or all doxastic laws of the category are preserved. As one kind of strictly conservative predicate change, we studied analytification, the change of a doxastic or typicality law into an analytical truth.

The status of analytic statements has been an important and controversial issue in the philosophy of the 20th century. The attack of Quine [23] on logical empiricism and especially on Carnap's philosophy [8, 9] is probably the best-known contribution to this discussion. Contrary to Quine, we assume that there are analytical laws. The notion arises from our formalism. In this respect, we subscribe to Carnap's ideas about intensionality. Our system actually implements a suggestion that we find in [10]. He claims that the intension of a predicate is formed by considering plausible and implausible possibilities. According to him, a linguist can determine the meaning of the German word "Pferd" by asking the native speaker Karl about implausible entities:

> All logically possible cases come into consideration for the determination of intensions. This includes also those cases that are causally impossible, i.e., excluded by the laws of nature holding in our universe, and certainly those that are excluded by laws which Karl believes to hold. Thus, if Karl believes that all P are Q by a law of nature, the linguist will still induce him to consider things that are P but not Q, and ask him whether or not he would apply to them the predicate under investigation (e.g., 'Pferd') [10, 38]

Our systems formalise exactly this notion of analyticity. Carnap emphasises the role of counterfactual entities to fix the meaning. Though analytical philosophers developed different views on meaning in the following decades, an affinity to highly counterfactual thought experiments – Putnam's examples of twin earth or brains in the vat are paradigmatic examples – raises the suspicion that they embraced the idea that meaning – the conceptual truth – is to be revealed in implausible cases.

Our discussion also provides arguments against emphasising analyticity. First, we have shown that analytic laws are not necessarily more stable than other laws but subject to changes. Second, these conceptual changes are motivated by our beliefs. In our view, analytic laws are not more fundamental than non-analytic ones. If predicate change occurs and if it is rational, analytic truth will finally rely on empirically grounded beliefs. Finally, if one is primarily interested in plausible entities, which we usually are, the question which laws qualify as analytic is of minor importance. In these respects, Quine's reservation against the notion of analyticity is justified. However, the late Carnap embraced the view that analyticity is often not important and vague. For a speaker like Karl, he notes, that "[t]his lack of clarity does not bother him much because it holds only for aspects which have very little practical importance for him" [10, 40].

Acknowledgments Open Access funding provided by Projekt DEAL. Work on this paper was funded by the DFG (Deutsche Forschungsgemeinschaft) within the project "The Structure of Representations in Language, Cognition, and Science" (Grant SFB991 D01). I am grateful to two anonymous reviewers, who provided many helpful suggestions.

Appendix

Proposition 3 *Any consistent analytic law in dynamic AD (and ADT / ADT 3) can be stipulated by inclusive and exclusive predicate changes.*

Proof We prove by induction, starting with atomic predicates and generalising to complex predicates.

- Base case: Analytic laws with one atomic predicate
  As said on page 27, the following two formulae are tautological:

  $$[\Phi \uparrow \Psi]\mathcal{A}\Psi\Phi$$
  $$[\Phi \downarrow \Psi]\mathcal{A}\Phi\Psi$$

  Therefore, if Φ is atomic, then the laws $\mathcal{A}\Psi\Phi$ and $\mathcal{A}\Phi\Psi$ can be generated. $\Phi \uparrow \Psi$ can be called an inclusive change (of Φ) towards $\mathcal{A}\Psi\Phi$ and $\Phi \downarrow \Psi$ an exclusive change (of Φ) towards $\mathcal{A}\Phi\Psi$.
- Induction step
  Inclusive and exclusive change can be generalised to arbitrary complex predicates Ξ

  1. Inclusive predicate change towards $\mathcal{A}\Psi\Xi$.

     (a) If Ξ is atomic, include by $\Xi \uparrow \Psi$.
     (b) For $\Xi = \Phi_1 \cap \Phi_2$, make an inclusive predicate change towards $\mathcal{A}\Psi\Phi_1$ *and* an inclusive predicate change towards $\mathcal{A}\Psi\Phi_2$.
     (c) For $\Xi = \Phi_1 \cup \Phi_2$, make an inclusive predicate change towards $\mathcal{A}\Psi\Phi_1$ *or* an inclusive predicate change towards $\mathcal{A}\Psi\Phi_2$.
     (d) For $\Xi = -\Phi$, make an exclusive predicate change of Φ towards $\mathcal{A}\Phi - \Psi$, which is equivalent to $\mathcal{A}\Psi - \Phi$.

  2. Exclusive predicate change towards $\mathcal{A}\Xi\Psi$.

     (a) If Ξ is atomic, exclude by $\Xi \downarrow \Psi$.
     (b) For $\Xi = \Phi_1 \cup \Phi_2$, make an exclusive predicate change towards $\mathcal{A}\Phi_1\Psi$ *and* an exclusive predicate change towards $\mathcal{A}\Phi_2\Psi$.

(c) For $\Xi = \Phi_1 \cap \Phi_2$, make an exclusive predicate change towards $\mathcal{A}\Phi_1\Psi$ *or* an exclusive predicate change towards $\mathcal{A}\Phi_2\Xi$.
(d) For $\Xi = -\Phi$, make an inclusive predicate change of Φ towards $\mathcal{A} - \Psi\Phi$, which is equivalent to $\mathcal{A} - \Phi\Psi$.

Thus, any (consistent) analytic statements (no matter how complex) can be generated by a series of predicate changes. □

References

1. Alchourron, C.E., Gärdenfors, P., Makinson, D. (1985). On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, *50*(2), 510–530.
2. Baltag, A., & Renne, B. (2016). Dynamic epistemic logic. In Zalta, E.N. (Ed.) *The stanford encyclopedia of philosophy, winter 2016 edn, metaphysics research lab*. Stanford: stanford university.
3. Baltag, A., Moss, L.S., Solecki S. (1998). The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th conference on Theoretical aspects of rationality and knowledge* (pp. 43-56). Burlington: Morgan Kaufmann Publishers.
4. van Benthem, J. (2004). Belief revision and dynamic logic. *Journal of Applied Non-Classical Logics*, *14*(2), 1–25.
5. van Benthem, J. (2011). *Logical dynamics of information and interaction*. Cambridge: Cambridge University Press.
6. van Benthem, J., van Eijck, J., Kooi, B. (2006). Logics of communication and change. *Information and Computation*, *204*(11), 1620–1662.
7. Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.
8. Carnap, R. (1928). Der logische Aufbau der Welt. Felix Meiner, Berlin.
9. Carnap, R. (1947). *Meaning and necessity: a study in semantics and modal logic*. Chicago: University of Chicago Press.
10. Carnap, R. (1955). Meaning and synonymy in natural languages. *Philosophical studies*, *6*(3), 33–47.
11. van Ditmarsch, H., van der Hoek, W., Kooi, B. (2007). *Dynamic epistemic logic*. Dordrecht: Springer.
12. Fleck, L. (1979). *Genesis and development of a scientific fact*. Chicago: University of Chicago Press.
13. Hall, B.K. (1999). The paradoxical platypus. *BicoScience*, *49*(3), 211–218.
14. Hampton, J.A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, *15*(1), 55–71.
15. Kraus, S., Lehmann, D., Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, *44*(1-2), 167–207.
16. Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
17. Lewis, C.I. (1943). The modes of meaning. *Philosophy and Phenomenological Research*, *4*(2), 236–250.
18. Lewis, D. (1969). *Convention. a philosophical study*. Cambridge: Harvard University Press.
19. Lewis, D. (1973). *Counterfactuals*. Cambridge: Harvard University Press.
20. Linné, C. (1735). Systema naturae 1. http://www.biodiversitylibrary.org.
21. Linné, C. (1758). Systema naturae 10, pt. 1.http://www.biodiversitylibrary.org,.
22. Makinson, D., & Hawthorne, J. (2015). In Koslow, A., & Buchsbaum, A. (Eds.) *Lossy inference rules and their bounds: a brief review*, (pp. 385-407). Birkhäuser: Cham.
23. Quine, W.V. (1951). Two dogmas of empiricism. *The Philosophical Review*, *60*(1), 20–43.
24. Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*, 328–350.
25. Rosch, E., & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.
26. Rott, H. (2009). Shifting priorities: Simple representations for twenty-seven iterated theory change operators. In Makinson, D., Malinowski, J., Wansing, H. (Eds.) Towards mathematical philosophy (pp. 269-296). Dordrecht: Springer.
27. Schurz, G. (2001). What is 'normal'? An evolution-theoretic foundation for normic laws and their relation to statistical normality. *Philosophy of Science*, *68*, 476–497.

28. Schurz, G. (2012). Prototypes and their composition from an evolutionary point of view. In Hinzen w, & Machery, E. (Eds.) (pp. 530-553). Oxford: The Oxford Handbook of Compositionality, Oxford University Press.
29. Strößner, C. (2015). Normality and majority: Towards a statistical understanding of normality statements. *Erkenntnis*, *80*(4), 793–809.
30. Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.
31. Veltman, F. (1996). Defaults in update semantics. *Journal of philosophical logic*, *25*(3), 221–261.