Model Theoretic Explanations
in the Theory of Dense Linear Orderings

Justin Bledin

January 2, 2007

1 Mathematical Explanation

There is now a general consensus among philosophers of mathematics that mathematical explanations exist. Work by Mancosu, Sandborg, Hafner and others over the past decade (see Mancosu (2001) for a survey of recent research in this area) has evidenced a strong concern with the problem of explanation in both the mathematical community and in a long line of philosophers dating back to Aristotle. Even so, explicating exactly what a 'mathematical explanation' actually is has proved elusive. In contrast with the philosophy of science, where strong causal and unification traditions strive to unravel the nature of scientific explanation, discussions of mathematical explanation remain rather vague. This paper is part of a continued recent effort in the analytic literature to find a foothold in the fog.

Two oft-discussed contemporary accounts of mathematical explanation have been provided by Steiner (1978) and Kitcher (1984, 1989). Extending his work on scientific explanation, Kitcher's global account takes explanation as theoretical unification, or the linking of mathematical ideas. Steiner rather attempts to draw a boundary between explanatory and non-explanatory proofs and provides a framework for distinguishing between them. For Kitcher, a whole mathematical theory or system explains; for Steiner, explanatoriness is a property of specific mathematical proofs. While problems with Kitcher's position have been discussed elsewhere (see Mancosu & Hafner (2005a) for an application of Kitcher's unification account to real algebraic geometry), I am here concerned with testing the applicability of Steiner's local conception of explanation to actual practice. In this study from model theory (a branch of mathematical logic studying abstract structures), Steiner's theory is brought to bear on three alternative proofs of a central result in the theory of dense linear orderings. I argue that, on Steiner's account, all three proofs explain, yet restricting ourselves to Steiner's criteria for explanation ignores salient differences in how the proofs are able to explain the same mathematical fact. This suggests the need for a broader account of mathematical explanation incorporating various modes of structural/syntactic argumentation.

2 The ω-Categoricity of Dense Linear Orderings

So as not to obscure the central claims of this paper, I will try to simplify or omit technical material wherever possible, presenting highly informal proofs (though in each example, it should still be clear from the description how a more formal argument would run). Nonetheless, while the theorem to be proved is elementary, the variety of proof techniques presented below requires some work in introducing various model-theoretic methods.

Formally, a dense linear ordering without endpoints (in what follows, I use 'DLO' for both the theory of dense linear orderings without endpoints and the structures themselves; the use should be clear from context) is a set of elements satisfying the following axioms:

(i) $\forall x \neg(x < x)$

(ii) $\forall xy(x = y \vee x < y \vee y < x)$

(iii) $\forall xyz(x < y \wedge y < x \rightarrow x < z)$

(iv) $\forall xy(x < y \rightarrow \exists z(x < z \wedge z < y))$

(v) $\forall x \exists z(z < x)$, $\forall x \exists z(x < z)$.

The conditions (i)-(iii) give us a linear ordering, (iv) ensures the ordered set is dense (*i.e.*, between any two elements in the ordering, there is an intermediate element), (v) ensures the ordering has no endpoints. The axioms (iv) and (v) - the 'extension axioms' - will play a key role in what follows. Examples of orderings satisfying (i)-(iii) and (v) are the ordered integers $\langle \mathbb{Z}, < \rangle$, satisfying (i)-(v) are the countable ordered rationals $\langle \mathbb{Q}, < \rangle$ and satisfying (i)-(iv) only are the ordered rationals taken over a closed interval.

The main theorem under investigation is easily stated:

Theorem 1 *DLO is ω-categorical*

In other words, the theory of DLO has a unique countable model (up to isomorphism). The term 'ω-categoricity' derives from the wider model-theoretic project to classify the models of particular theories (up to isomorphism) based on their cardinality, or size. If two models are isomorphic, they are essentially the same structure. Though the models may have different domains and the constants, relations and functions defined over these domains may have differing interpretations, each model can be mapped onto the other preserving all structural features:

Definition 1 *Let $A = \langle \eta, <^\eta \rangle$ and $B = \langle \xi, <^\xi \rangle$ be linear orderings. An isomorphism is a one-to-one mapping f: $A \mapsto B$ such that $x <^\eta y \leftrightarrow fx <^\xi fy$.*

This definition will be particularly important in the discussion of back-and-forth equivalence. But without further ado, we are ready for the proofs.

2.1 Back-and-Forth Equivalence

Before presenting a technical syntactic argument involving quantifier elimination in the next section, I begin with a less round-about, purely structural proof that DLO is ω-categorical. Indeed, the original proof of this result allegedly goes back to Cantor who used back-and-forth methods to show that the elements of any two countable dense linear orderings without endpoints could be mapped from one structure to the other and back again. The modern incarnation of the back-and-forth argument is Fraïsse's notion of 'back-and-forth equivalence', introduced in the 1950s and playfully presented here, following Hodges (1997: 74-81), in terms of Ehrenfeucht-Fraïsse games.

The players are ∀belard and ∃loise, a twelfth-century Parisian logician and the niece of a Notre Dame canon. Given two linear orderings $A = \langle \eta, <^\eta \rangle$ and $B = \langle \xi, <^\xi \rangle$, ∀belard wants to prove A different from B while ∃loise wants to prove them identical. They take turns choosing elements a_i from A and b_i from B at each step of the game, ∀belard freely choosing from either A or B and ∃loise choosing from the opposite structure. At the end of the game with a (countably) infinite number of steps, denoted $EF_\omega(A, B)$, sequences of elements $a = \{a_i\}$ and $b = \{b_i\}$ have been chosen from A and B respectively with the pair (a, b) known as the final 'play'. The play (a, b) is a win for ∃loise if there exists an isomorphism $f : \langle a \rangle_A \mapsto \langle b \rangle_B$ between substructures $\langle a \rangle_A$ and $\langle b \rangle_B$ of A,B such that $fa = b$. In other words, the substructures of A and B generated from the elements $\{a_i\}$ and $\{b_i\}$ must be structurally equivalent.

In this context, Fraïsse's 'back-and-forth equivalence' can now be defined as follows:

Definition 2 *Two linear orderings A and B are back-and-forth equivalent, $A \sim_\omega B$, if ∃loise can always win the game $EF_\omega(A, B)$.*

When ∃loise knows of an isomorphism $f : A \mapsto B$, $EF_\omega(A, B)$ is easily won as she can choose $f(a_i)$ from B when ∀belard chooses the corresponding element a_i from A and $f^{-1}(b_i)$ when ∀belard picks b_i. The game is more interesting when ∃loise knows of no such isomorphism but fortunately there is a useful criterion for determining exactly when two structures are back-and-forth equivalent:

Definition 3 *A back-and-forth system from A to B is a set of functions J satisfying the following conditions:*

(BF1) each $f \in J$ is an isomorphism $f : \langle a \rangle_A \mapsto \langle b \rangle_B$

(BF2) J is non-empty

(BF3) $\forall f \in J$ and $c \in A$, there is $g \supseteq f$ such that $g \in J$ and $c \in dom(g)$

(BF4) $\forall f \in J$ and $d \in B$, there is $g \supseteq f$ such that $g \in J$ and $d \in im(g)$

Theorem 2 *A,B are back-and-forth equivalent $\leftrightarrow$ there is a back-and-forth system from A to B.*

The link between back-and-forth equivalence and back-and-forth systems should hopefully appear somewhat intuitive. To repeat, at each move in $EF_\omega(A, B)$, ∀belard is busy choosing elements from the linear orderings trying to back ∃loise into a corner where she can no longer choose appropriate corresponding elements in the opposite structures. If ∃loise is playing well, consistently finding matching elements to ∀belard's choices, then a snapshot of the game at any given moment would reveal a partial play (a', b') such that there exists an isomorphism $f : \langle a' \rangle_A \mapsto \langle b' \rangle_B$ with $fa' = b'$. Such an isomorphism must exist as if the suborderings $\langle a' \rangle_A$ and $\langle b' \rangle_B$ are structurally different, so too are the full orderings A and B, precluding ∃loise from winning $EF_\omega(A, B)$ (so A and B are not back-and-forth equivalent).

A back-and-forth system ensures that at *any* moment in the countable game $EF_\omega(A, B)$, ∃loise is on pace to win. (BF1) and (BF2) tell us that ∃loise is not doomed from the very beginning; (BF3), the 'forth' step, states that at each incremental step in the game, if ∀belard chooses any element from A, ∃loise can choose a corresponding structurally equivalent element from B; (BF4), the 'back' step, states that if ∀belard chooses any element from B, ∃loise can respond by picking an appropriate element from A. Putting this all together, the presence of a back-and-forth system between the orderings A and B ensures ∃loise has a winning strategy for $EF_\omega(A, B)$.

Given back-and-forth systems as criteria for back-and-forth equivalence, perhaps the reader has already anticipated the punch-line:

Theorem 3 *Let A,B be countable linear orderings. Then A,B are isomorphic ↔ they are back-and-forth equivalent.*

The argument in the (←) direction is essentially an inductive one (the other direction trivially holds as already indicated above). If A,B are back-and-forth equivalent, there exist isomorphic substructures $\langle a \rangle_A$ and $\langle b \rangle_B$ of A and B and at each incremental step we can expand the substructures one element and maintain their structural equivalence. After countably many steps, we are left with a full-blown isomorphism between A and B.

To prove the ω-categoricity of DLO, it must now only be shown that for countable dense linear orderings $A = \langle \eta, <^\eta \rangle$ and $B = \langle \xi, <^\xi \rangle$ without endpoints, a back-and-forth system exists between them:

Theorem 4 *If A,B are countable dense linear orderings without endpoints, there exists a back-and-forth system from A to B.*

Proof : In the base case, where $A = \{a\}$ and $B = \{b\}$ are one element orderings, let $f(a) = b$; so (BF2) holds. Now assume $\langle a \rangle_A = a_1 < ... < a_n$, $\langle b \rangle_B = b_1 < ... < b_n$ and there exists an isomorphism $f : \langle a \rangle_A \mapsto \langle b \rangle_B$ with $f(a) = b$. We must show that the 'forth' step holds. Choose $c \in A$ not in $\{a_1, ..., a_n\}$. Either $c < a_i$, $a_i < c \; \forall a_i$, or $a_i < c < a_j$ for some $a_i, a_j \in \{a_1, ..., a_n\}$. In the latter case, the density of B ensures we can find some extension $g \supseteq f$ and $g(c) = d \in B$ such that $g(a_i) < d < g(a_j)$. In the previous cases, the infinite extension of B in both directions ensures we can find d. An analogous argument works for the 'back' step.

Note that in this proof, the extension axioms (iv) and (v) are doing all the work. No matter which element ∀belard chooses from either structure, since *both* A and B are DLO, ∃loise can always match it.

2.2 Quantifier Elimination

The second proof takes us back to the early days of model theory, the method of quantifier elimination originating in Tarski's Warsaw seminar in the late 1920s. The idea is simple but useful: show that relative to a theory T, all formulae in a first-order language L (including those with quantifiers) are equivalent to quantifier-free formulae. In the special cases where quantifier elimination is successful, the result is a condensed description of all complete extensions of T, simplifying the study of definable sets on models of T and usually leading to completeness and decidability proofs (see Hodges 1997: 60-1 for details).

The traditional approach to quantifier elimination (used here) differs sharply from the other model-theoretic proofs examined in this paper in that it is a heavily syntactic approach (model theorists such as Abraham Robinson later encouraged the use of good structural information about the models of T, rather than syntax, to show T admits quantifier elimination). Chang and Kiesler (1997: 49) write: "the method may be thought of as a direct attack on a theory." Though these 'attacks' are not very difficult, they can be quite tedious. To tighten the below proof, I follow Marker (2002: 66-7) in taking the completeness of DLO for granted at the onset. An alternative route, taken by Chang and Kiesler (50-4), is to prove DLO admits quantifier elimination and have completeness fall out as an easy consequence (the below proof combines elements of the proofs given in Chang and Kiesler (50-4) and Marker (66-7)).

Theorem 5 *DLO admits quantifier elimination*

Proof : It must be shown that for every formula ϕ in L, there exists a quantifier-free ψ such that DLO$\vdash \phi \leftrightarrow \psi$ (ϕ is equivalent to ψ mod DLO). First consider when ϕ is a sentence so as DLO is complete, either DLO$\vdash \phi$ or DLO$\vdash\sim \phi$. If DLO$\vdash \phi$, then DLO$\vdash \phi \leftrightarrow x_1 = x_1$; if DLO$\vdash\sim \phi$, then DLO$\vdash \phi \leftrightarrow x_1 < x_1$ (note that $x_1 = x_1$ and $x_1 < x_1$ could be replaced here by $\top$ and $\bot$ if our language allowed it).

So suppose ϕ is a formula with free variables $x_1, ..., x_n$. We show the set of atomic formulae Φ: $\{x_i = x_j, x_i < x_j\}$ forms an elimination set for the class of all models of DLO (*i.e.*, ϕ is DLO-equivalent to a Boolean combination of the formulae in Φ). Define an *arrangement* of $x_1, ..., x_n$ to be the finite conjunction of formulae $\Theta = \theta_1 \wedge \theta_2 \wedge ... \wedge \theta_n$ where each θ_i is of the form $x'_i < x'_{i+1}$ or $x'_i = x'_{i+1}$ for a renumbering of $x_1, ..., x_n$ as $x'_1, ..., x'_n$. Each arrangement gives us an exact ordering of the n free variables. Now when $n = 1$, an open formula ϕ is built from Φ so since DLO$\vdash x_1 = x_1$ and DLO$\nvdash x_1 < x_1$ we must have DLO$\vdash \phi$ (in which case DLO$\vdash \phi \leftrightarrow x_1 = x_1$) or DLO$\vdash\sim \phi$ (in which case DLO$\vdash \phi \leftrightarrow x_1 < x_1$). So consider when $n > 1$. If DLO$\cup\{\phi\}$ is inconsistent, DLO$\vdash \phi \leftrightarrow x_1 < x_1$. If DLO$\cup\{\phi\}$ is consistent with ϕ open, ϕ must be DLO-equivalent to a disjunction of finitely many arrangements of $x_1, ..., x_n$ as this exhausts what we can say about the variables in ϕ in terms of Boolean combinations of the atomic formulae in Φ (*i.e.*, DLO$\vdash \phi \leftrightarrow \bigvee \Theta$). As an example, the formula $\phi(x_1, x_2, x_3) = (x_1 < x_3) \wedge (x_2 < x_3)$ is equivalent (mod DLO) to $\bigvee \Theta = ((x_1 = x_2) \wedge (x_2 < x_3)) \vee ((x_1 < x_2) \wedge (x_2 < x_3)) \vee ((x_2 < x_1) \wedge (x_1 < x_3))$ as only these three possible arrangements of x_1, x_2, x_3 are compatible with ϕ.

We can now use this condition to show that *every* formula in L is DLO-equivalent to an open formula. Here, it suffices to show that if ϕ is an open formula, $\exists x_n \phi$ is DLO-equivalent to an open formula. By the above, DLO$\vdash \exists x_n \phi \leftrightarrow \exists x_n \bigvee \Theta$ so DLO$\vdash \exists x_n \phi \leftrightarrow \bigvee \exists x_n \Theta$. For each arrangment Θ, let Θ^* be the restricted arrangement of $x_1, ..., x_{n-1}$ obtained by omitting the literal involving x_n. Now the crucial step is this: given the extension axioms ((iv) and (v) above) for DLO, telling us that an element can be found in any possible interval over the ordering, $\exists x_n \Theta$ is equivalent (mod DLO) to Θ^*. Thus, DLO$\vdash \exists x_n \phi \leftrightarrow \bigvee \Theta^*$ where $\bigvee \Theta^*$ is a Boolean combination of the formulas in Φ so DLO admits quantifier elimination.

Let us take a moment to review the last step in the proof one more time. We are given an existence claim $\exists x_n \Theta$ about a particular element x_n in an ordering of $x_1, ..., x_n$. The crucial inference is that if $x_1, ..., x_{n-1}$ are ordered appropriately, the element x_n *must* exist. The extension axioms in the theory DLO tell us so: if $\exists x_n \Theta$ states that x_n is less (greater) than all of $x_1, ..., x_{n-1}$, such an element exists as DLO does not have endpoints; if $\exists x_n \Theta$ states that x_n lies between two elements among $x_1, ..., x_{n-1}$ (*i.e.*, $x_i < x_n < x_j$), such an element exists as DLO is dense. The upshot of all this is that if Θ^* holds, so does $\exists x_n \Theta$ and vice versa.

But we are not yet done. Now that DLO has been shown to admit quantifier elimination, there is the further step of showing DLO is ω-categorical. There are several ways to do this, all of them taking us into the more abstract realms of model theory. One nice link from quantifier elimination to ω-categoricity is provided by a theorem of Engler, Ryll-Nardzewski and Svenonius (Hodges: 171):

Theorem 6 *Let L be a countable first-order language and T a complete theory in L which has infinite models. Then the following are equivalent:*

(a) Any two countable models of T are isomorphic

(b) For each $x = (x_1, ..., x_n)$, there are only finitely many pairwise non-equivalent formulae $\phi(x)$ of L mod T

Given that DLO has the property of quantifier elimination, it is trivially shown that (b) is satisfied (note that this says there are only finitely many distinct possible arrangements of $x_1, ..., x_n$) so by (a), DLO is ω-categorical. I omit the proof of the Engeler, Ryll-Nardzewski and Svenonius theorem (which takes us into distant waters, omitting types and the like for those familiar with model theory).

2.3 Fraïsse Limits

The third proof puts formal languages aside, returning to pure structural analysis and introducing another of Fraïsse's contributions to model theory: the Fraïsse limit. Fraïsse's ingenious idea was that given a class of finite structures having various properties, we can amalgamate (or join) them together to form a 'limit' structure. In the case of linear orderings, the limit of the class of finite linear orderings is the ordered rationals $\langle \mathbb{Q}, < \rangle$. As Fraïsse limits are unique, the ω-categoricity of DLO follows immediately.

Unlike in the previous section, the focus here will be on giving a concise introduction to the Fraïsse limit concept - a truly remarkable way to view the rationals - rather than providing the skeleton of a proof (though again, it should still be clear how a formal proof would go). Most of the material here is found in Hodges (1997: 158-61). Formal proofs of the existence and uniqueness of Fraïsse limits can also be found in Hodges (161-4).

The starting point is a class K of finitely generated structures. With linear orderings, K is a collection of finite structures. K is called the *age* of some structure if K is non-empty and has the following properties:

- *Hereditary property* (HP): If $A \in K$ and B is a finitely generated substructure of A (*i.e.*, $B = \langle a \rangle_A$ for finite a), then B is isomorphic to some structure in K
- *Joint embedding property* (JEP): If A,$B \in K$, then there exists a $C \in K$ and embeddings $f : A \mapsto C$ and $g : B \mapsto C$ (an embedding from $A \mapsto C$ is an isomophism from A to some substructure of C).

In addition, the class of linear orderings has the following important property:

- *Amalgamation property* (AP): If A,B,$C \in K$ and $e : A \mapsto B$ and $f : A \mapsto C$ are embeddings, then there is a $D \in K$ and embeddings $g : B \mapsto D$ and $h : C \mapsto D$ such that $ge = hf$

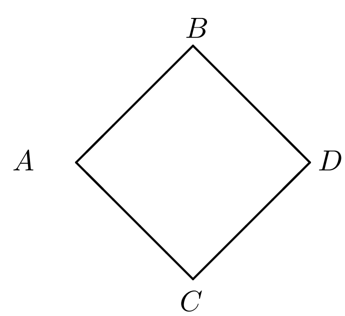

Figure 1: Amalgamation property (Hodges: 160)

As a trivial example, consider the ordering $B = 0 < 1 < 2 < 3 \in K$. As K is the class of all finite orderings, all substructures of B (such as $B' = 1 < 2 < 3$) belong to K (this is an instance of HP). Now let $C = 0 < 1/2 < 1$. C is clearly embeddable in B by letting $f(0, 1/2, 1) = (0, 1, 2)$, illustrating the JEP. But AP gives us more. As $A = 0 < 1$ is embeddable in both B and C, AP tells us that there exists some $D \in K$ and embeddings from $B \mapsto D$ and $C \mapsto D$ that are constant on A. Clearly, $D = 0 < 1/2 < 1 < 2 < 3$. Note that while HP and JEP ensure that we can fit every last structure in K together in a new larger structure, AP tells us that every element in every structure in K ultimately appears in these larger structures. By the AP, the countable limit of the class K of finite orderings must then be a dense linear ordering without endpoints so K tends to the rationals rather than, say, the integers or natural numbers (though K is still the age of both $\langle \mathbb{Q}, < \rangle$ and $\langle \mathbb{Z}, < \rangle$).

This brings us to Fraïsse's Theorem (Hodges: 160):

Theorem 7 *Let L be a countable language, K a non-empty finite/countable collection of finitely generated L-structures which has HP, JEP and AP. Then there is a unique (up to isomorphism) L-structure D of cardinality $\leq \omega$ such that K is the age of D and D is homogeneous.*

Definition 4 *A structure D is homogeneous if every isomorphism between finitely generated substructures of D extends to an isomorphism from $D \mapsto D$ (*i.e.*, an automorphism of D).*

The structure D in Fraïsse's Theorem is what I, following Hodges, have been calling the Fraïsse limit (it is also known as the 'universal homogeneous structure of age K'). As already mentioned, when K is the class of finite orderings, D is the ordered rationals. Fraïsse's Theorem thus says that DLO is ω-categorical. Other interesting examples of Fraïsse limits are the countable atomless Boolean algebra (where K is the class of finite Boolean algebras) and the celebrated 'random graph' (where K is the class of all finite graphs).

But what exactly does it mean to say $\langle \mathbb{Q}, < \rangle$ is 'homogeneous'? Well if every isomorphism between substructures of the ordered rationals extends to an automorphism, then when we take a peek at some local region of $\langle \mathbb{Q}, < \rangle$ and it looks the same, it really is. Again, the extension axioms play the central role as the 'homogeneity' of the rationals is precisely their density and lack of endpoints. By contrast, consider an ordering of the natural numbers $\langle \mathbb{N}, < \rangle$. As only the zero element is a lower bound of all other elements and the naturals are not dense (this creates a problem as the substructures $A' = 0 < 1$ and $A'' = 0 < 2$ are isomorphic but any isomorphism between them can clearly not be extended to an automorphism of $\langle \mathbb{N}, < \rangle$), the ordered naturals are not homogeneous. 'Homogeneity', then, appears to be a *characterizing property* of the ordered rationals among the family of linear orderings of size $\leq \omega$. But we are getting slightly ahead of ourselves and having finished our preparations, it is time to explore Steiner's account of mathematical explanation.

3 Testing Steiner's Model

In this section, Steiner's criteria for explanation are applied to the three model-theoretic proofs. As Steiner's account has been discussed extensively in Resnik and Kushner (1987) and Hafner and Mancosu (2005a), the presentation of the actual model will be brief. The focus is on how Steiner's account can be suitably tested against examples from actual mathematical practice. While other authors have questioned the applicability of Steiner's model (more on this below), I will argue that, on Steiner's account, all three proofs of the ω-categoricity of DLO explain. However, far from being a virtue of Steiner's approach (if one does, like me, believe such proofs *do* explain, for whatever that's worth), the inability for Steiner's explanation criteria to distinguish between the three alternative model-theoretic methods indicates something is missing in Steiner's model. In section 4, I argue that a broader, more heterogeneous view of mathematical explanation incorporating various modes of structural/syntactic reasoning would be needed to fill this gap.

The goal of Steiner's (1977) local approach to mathematical explanation is a clean divide between proofs that explain and those that do not. For Steiner, the distinctive feature of explanatory proofs is their dependence on a 'characterizing property' of a mathematical object, a property "unique to a given entity or structure within a family or domain of such entities or structures," (Steiner: 143) where 'family' is taken as primitive. Steiner does acknowledge some ambiguity in the notion of a characterizing property as a mathematical object may be part of multiple families or have several characterizing properties even in the context of a single domain. However, there must be a clear link between the characterizing features of object(s) appearing in a mathematical theorem and the proof that explains this theorem. Steiner writes:

> "My proposal is that an explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that from the proof it is evident that the result depends on the property. It must be evident, that is, that if we substitute in the proof a different object of the domain, the theorem collapses; more, we should be able to see as we vary the object how the theorem changes in response." (143)

A second criterion for explanation enters here as the explanatory proof must also be generalizable, or 'deformable'. By replacing mathematical objects in a proof with other entities or structures, we obtain new proofs and theorems dependent on how the relevant characterizing properties of the old and new entities/structures differ. Explanation is ultimately a relation between a series of proofs and theorems, rooted in the characterizing properties of the mathematical objects that inhabit them.

Applying Steiner's theory (*i.e.*, establishing whether a particular mathematical proof is explanatory) thus requires the following steps:

(S1) identify the mathematical entities or structures mentioned in the theorem and define the family to which they belong;

(S2) isolate the characterizing properties of the objects from (S1) on which the proof depends;

(S3) demonstrate that the proof no longer works if other entities or structures from the same family with differing characterizing properties are substituted in the proof;

(S4) show that by 'deforming' the proof (by varying characterizing properties), we can obtain proofs of related theorems.

The steps (S1) and (S2) are prerequisites - with no characterizing properties, Steiner's account cannot even get off the ground; (S3) is what Hafner and Mancosu call the 'dependence test' (2005a: 234); (S4) expresses Steiner's claim that "it is not, then, the general proof which explains; it is the *generalizable* proof." (1977: 144)

At first glance, (S1) and (S2) are seemingly innocuous. Proofs discussed by Steiner (144-6) of the irrationality of $\sqrt{2}$ (using the unique prime power expansion of 2), the Pythagorean theorem (requiring that right triangles can be decomposed into two similar triangles that are also similar to the whole), the sum of the first n integers (using symmetric and geometric properties of the sum

$1 + 2 + ... + n$) and Euler's identity (proceeding from enumeration properties of infinite sums and products) suggest that mathematical entities and their characterizing properties abound in mathematical proofs. Yet this may not be the case. Resnik and Kushner (1987: 147-150) provide two counterexamples, a proof of the intermediate value theorem in analysis and Henkin's completeness proof of first-order logic, where it is unclear which characterizing properties are involved. In their case study of Pringsheim's work in the theory of infinite series, Hafner and Mancosu (2005a: 230-4) find themselves in a paradoxical situtation trying to find a characterizing property of a completely arbitrary sequence. Whether these counterexamples can be turned around or not, they indicate some immediate difficulties in the general applicability of Steiner's model to mathematical practice.

Fortunately for us, the benefits of choosing a model-theoretic example pay off handsomely here. (S1) and (S2) are easily implemented: the mathematical structure under consideration is a countable dense linear ordering without endpoints. DLO belongs to the family of all linear orderings. The three proofs aim to show DLO is unique (up to isomorphism) and this is clearly stated in the theorem that 'DLO is ω-categorical'. The characterizing properties of DLO on which all three proofs crucially depend are precisely its countability, density and lack of endpoints (expressed by the extension axioms) or alternatively, heeding our earlier discussion of Fraïsse limits, its homogeneity.

But what is most interesting is the distinct ways in which the proofs refer to and require (satisfying the dependence test (S3)) the characterizing properties of DLO. In the back-and-forth equivalence proof, recall that ∀belard is trying to show that two linear orderings are different while ∃loise is busy trying to pick elements from the two structures that counterbalance those chosen by ∀belard. Games aside, we are incrementally building isomorphisms from larger and larger finitely generated substructures of the orderings hoping that, after countably many steps, the union of these partial maps is a full isomorphism between the structures. As mentioned, the extension axioms do all the work. The density and lack of endpoints of either structure ensure that no matter which elements ∀belard chooses, ∃loise has a winning strategy. Put differently, the characterizing properties of DLO ensure that we can build a back-and-forth system between any two countable dense linear orderings without endpoints and DLO is therefore ω-categorical.

In the argument from quantifier elimination, the extension axioms are also essential for the proof to go through. The main step of the proof is showing the equivalence (mod DLO) of the formulae $\exists x_n \Theta$ and Θ^*, *i.e.*, that a variable x_n can be found in a particular place in a particular arrangement of $x_1, ..., x_{n-1}$. Given the characterizing properties of DLO, the existence claim $\exists x_n \Theta$ must hold so long as $x_1, ..., x_{n-1}$ are ordered appropriately. We thus conclude that DLO$\vdash \exists x_n \Theta \leftrightarrow \Theta^*$ and it is a short step further to show DLO has quantifier elimination. By invoking the theorem of Engler, Ryll-Nardzewski and Svenonius, DLO is then shown to be ω-categorical.

In the proof using Fraïsse limits, the role of the characterizing properties of DLO is even more pronounced (the countability of DLO is particularly important here) as they allow for the characterization of DLO as the Fraïsse limit of the class of finite orderings. As this class has the properties HP, JEP and AP, all finite orderings can be amalgamated together to form a countable homogeneous structure, the densely ordered rationals $\langle \mathbb{Q}, < \rangle$. The ω-categoricity of DLO then follows immediately from the rationals' role as the unique Fraïsse limit of the collection of finite orderings.

It is easily shown that the latter two proofs no longer work if we replace DLO with other structures in the family of linear orderings having different characterizing properties (*i.e.*, that are finite/uncountable or for which the extension axioms do not hold). This is straightforward in the case of Fraïsse limits as countability and homogeneity are defining features of the limit. Though the class of finite orderings is the age of both the ordered rationals $\langle \mathbb{Q}, < \rangle$ and integers $\langle \mathbb{Z}, < \rangle$, the amalgamation property AP ensures that the finite orderings tend to only the dense rationals. The equivalence of $\exists x_n \Theta$ and Θ^* in the quantifier elimination proof also fails if we consider the theory T of linear orderings without the extension axioms or the theory of dense linear orderings *with* endpoints. Given the arrangement Θ^*, there is no longer any guarantee that x_n can be found in a particular location in the ordering of $x_1, ..., x_{n-1}$. In this case, T$\nvdash \exists x_n \Theta \leftrightarrow \Theta^*$ so the theory does not admit quantifier elimination (note where T = theory of dense orderings with endpoints, all formulae in L can still be shown T-equivalent to an elimination set Φ containing the atomic formulae *and* the formula $\forall x \exists z (z < x)$ and/or $\forall x \exists z (x < z)$).

The situation is slightly more complicated in the case of back-and-forth equivalence as back-and-forth methods can be used to show a variety of linear orderings are isomorphic, not just endpoint-free dense ones. Restricting our attention to the countable case, a back-and-forth system can be constructed between two orderings A and B so long as the same extension axioms are satisfied by both structures. To illustrate a case where back-and-forth equivalence (and hence isomorphism) breaks down, consider a back-and-forth game where A has an endpoint while B does not. ∀belard can then win the game by playing the following strategy: first choose the endpoint $a \in A$ (without loss of generality, assume this is a right endpoint) and then after ∃loise has chosen a corresponding element $b \in B$, choose any element in B to the right of b. As there are no elements in A to the right of the endpoint a, ∃loise has lost the game. In other words, there is no back-and-forth system between A and B so the structures are not isomorphic. By analogous reasoning, it can also be shown that dense and non-dense countable orderings are not isomorphic. Thus, as opposed to the previous proofs where the theorem collapses if *any* other ordering besides DLO is substituted in the proof, the argument from back-and-forth equivalence only fails when the structures A,B compared in the proof have differing characterizing properties.

With (S1), (S2) and (S3) now satisfied, (S4) can also be implemented, *i.e.*, the proofs are generalizable by varying the characterizing properties of the structures mentioned in them. By applying the concept of a Fraïsse limit to other families of structures besides linear orderings, we obtain similar results: looking at the collection of all finitely generated Boolean algebras, the unique Fraïsse limit is the countable atomless Boolean algebra whose 'homogeneity' is precisely its property of having no non-zero minimal elements (the theory of atomless Boolean algebras consists of the axioms for distributive lattices together with the extension axiom $\forall x (x \neq y \rightarrow \exists y (0 < y \wedge y < x)))$; for the class of all finite graphs, the unique Fraïsse limit is the 'random graph' with the interesting property that for any two finite disjoint sets X and Y of vertices in the random graph, there is an element z$\notin X \cup Y$ connected (by an edge) to every vertex in X and no vertex in Y (the theory of the random graph consists of the regular graph axioms together with the extension axiom $\forall x_1 ... x_n \forall y_1 ... y_n (\bigwedge \bigwedge x_i \neq y_j \rightarrow \exists z (\bigwedge (R(x_i, z) \wedge \sim R(y_i, z)))))$. Interestingly, the random graph can be used to study the asymptotic properties of random finite graphs, bridging graph theory with probability theory (see Marker 2002: 45-6 for a nice presentation of the remarkable Zero-One Law for Graphs). But for our present purposes, the main point here is that by holding the proof-idea (Fraïsse limits) constant and substituting other structures for DLO with differing characterizing properties (though they are all necessarily extension properties), we obtain the new related theorems 'the theory of atomless Boolean algebra is ω-categorical' and 'the theory of the random graph is ω-categorical'.

The method of quantifier elimination has also been successfully applied to other theories besides DLO, including the theory of Abelian groups, Presburger Arithmetic and the theories of real and algebraically closed fields (RCF/ACF). A high point for model theory, Tarski (1948) used quantifier elimination to show the decidability of RCF. Unfortunately, these examples do not really help our case as to satisfy (S4), we are looking for examples of theories where quantifier elimination is central to a proof that any countable models of the theory are isomorphic (recall that in addition to the proof that DLO has quantifier elimination, the move to ω-categoricity still required the Engler, Ryll-Nardzewski and Svenonius theorem). Fortunately such examples do exist: the theory of atomless Boolean algebras, the theory of infinite Abelian groups with finite exponent, the theory of an equivalence relation with infinitely many infinite classes, to name a few. By applying the original proof strategy to these theories (whose models have differing characterizing properties), we can show the theories are ω-categorical. So though quantifier elimination still does not apply to the majority of interesting mathematical theories and not all theories admitting quantifier elimination are ω-categorical, these examples are sufficient to ground (S4), indicating how we can obtain related theorems through deformation of the quantifier elimination proof.

Of the three proofs, the argument from back-and-forth equivalence is arguably the most deformable as back-and-forth techniques can be used to show isomorphism both within the family of linear orderings and in other classes of structures. As indicated earlier, only slight changes to the proof given in this paper result in new proofs that the theory of dense linear orderings with two endpoints is ω-categorical, the theory of dense linear orderings with right endpoints is ω-categorical, and so on. Other deformations of the back-and-forth proof also show that more diverse structures are isomorphic. Here is a concrete example (adapted from Hodges: 79-80):

Theorem 8 *The theory of atomless Boolean algebras is ω-categorical*

Proof : In the base case, where A and B are Boolean algebras with only zero and one elements, let $f(0_A) = 0_B$ and $f(1_A) = 1_B$; so (BF2) holds. Now consider $\langle a \rangle_A$ with atoms $a = a_1, ..., a_n$ and $\langle b \rangle_B$ with atoms $b_1, ..., b_n$ and assume there exists an isomorphism $f : \langle a \rangle_A \mapsto \langle b \rangle_B$ with $f(a) = b$. We must show that the 'forth' step holds. Choose $c \in A$ not in $\{a_1, ..., a_n\}$. Then the structural identity (what Hodges calls 'isomorphism type') of c over $\langle a \rangle_A$ is determined by, for each atom a_i, whether c$\wedge a_i$ equals a_i, 0 or neither. c$\wedge a_i$ is the greatest lower bound of c and a_i so, intuitively, the split concerns whether c lies above a_i, beside a_i or below a_i. Now as B is atomless, we can find some extension $g \supseteq f$ and $g(c) = d \in B$ such that $d \wedge g(a_i) = g(a_i), 0$ or neither $\leftrightarrow c \wedge a_i = a_i, 0$ or neither. An analogous argument works for the 'back' step.

As with the DLO proof, the extension axiom does all the work, though this time the extension axiom concerns the absence of non-zero generators (or atoms) in certain special Boolean algebras, a property closely resembling but distinct from the density of certain special linear orderings. But while the characterizing properties on which the proofs depend have shifted, the proof-idea remains unaltered.

Let us now summarize our findings regarding the three alternative model-theoretic proofs that all countable dense linear orderings are isomorphic: each proof refers to and crucially depends on the countability and homogeneity of DLO; if other linear orderings are substituted in the proof, the theorem collapses (with some exceptions in the back-and-forth case); nonetheless, by examining structures in other classes with similar but distinct characteristic properties (reflected in the extension axioms in the theories of which these structures are models) while holding the various proof-ideas constant, we obtain new proofs of closely related theorems, such as the uniqueness (up to isomorphism) of countable atomless Boolean algebras. In short, as all three model theoretic proofs satisfy (S1)-(S4), I conclude that, on Steiner's account, all three proofs *explain* why DLO is ω-categorical.

Before reflecting on this further, one more issue needs to be cleared up. The reader with some prior familiarity with model theory may wonder whether we are warranted in even calling the above examples three distinct proofs at all. For to be sure, there is much overlap between the various model theoretic methods not explicitly mentioned in the presentation above: the proof of the uniqueness of Fraïsse limits, crucial for linking the discussion of Fraïsse limits to the ω-categoricity of DLO, is essentially a back-and-forth argument; a theory T has quantifier elimination if and only if $T_\forall$, the set of universal consequences of T, has the amalgamation property (see Hodges: 203 for proof); Ehrenfeucht-Fraïsse games can be used to find elimination sets (Hodges: 85-9). In this light, the boundaries of the proofs seem somewhat arbitrary. To some extent, I agree but have no desire to start building perimeter fences. I think there is enough variation in the three model-theoretic arguments that it should be clear that they contribute to proving the same mathematical fact in distinct and interesting ways, regardless of whether borders between them blur.

4 Structural and Syntactic Methods

To avoid misunderstandings of my argument, it must be stated at the onset that I do *not* intend in this section to give my own definitive account of mathematical explanation, though hints of such an account may present themselves. I do *not* argue that Steiner's account of explanation should be accepted under the proviso that further work be done to clarify exactly how proofs depend on the characterizing properties of the mathematical objects that appear therein. Rather, I argue that *even if* we accept Steiner's model, it still does not do the necessary work (I take a more positive line on mathematical explanation in the final remarks in section 5). Resnik/Kushner and Hafner/Mancosu provide examples of proofs that explain but fail to meet Steiner's criteria to show "Steiner's account seems bound to *undergenerate*, i.e., it seems thus blocked from fully capturing the intuitive notion of explanatory proof operative in mathematical practice." (Hafner and Mancosu 2005a: 237) I claim that Steiner's account undergenerates in another important way: though all three of our model-theoretic proofs meet Steiner's explanation criteria, each proof explains the ω-categoricity of dense linear orderings in its own distinct way and an account of explanation that simply seeks to identify characterizing properties ignores the complex and interesting roles such properties play in mathematicians' attempts to convince others of this or that mathematical claim (a reminder here that I am working under the assumption that these proof explain precisely because they meet Steiner's criteria). Steiner's model seems thus blocked from fully capturing the variety of *explanatory methods* operative in mathematical practice.

So how exactly does Steiner's account fall short? Answering this question requires a closer look at how the model-theoretic proofs in this paper explain. The back-and-forth argument explains by introducing a new equivalence relation between two structures A and B. Like the notion of an isomorphism, the relation of back-and-forth equivalence is a structural one. Yet ultimately the focus is not so much on the full structures A and B themselves but on the individual elements that comprise the structures. If A,B are back-and-forth equivalent, each element in either A or B can be mapped to a corresponding structurally equivalent element in the opposite structure and this step can be implemented countably many times. The notion of back-and-forth equivalence thus provides a local view of structural identity. Harnessing this local view, we come to understand that DLO is ω-categorical by a ground-level inductive argument: as we can get isomorphisms to work between incrementally larger subsets of elements in dense linear orderings (up to ω), we can get an isomorphism to work between the full structures.

The Fraïsse limit proof also explains the uniqueness (up to isomorphism) of DLO by introducing a new structural concept. Fraïsse taught us that in certain classes of structures, there exists a special universal structure, the countable limit obtained by amalgamating all the structures in the class. Such structures have the nice property of homogeneity, a strong symmetry property that manifests itself in different forms in different Fraïsse limits, from the unusual connectivity of the random graph, the density near zero of the countable atomless Boolean algebra and in the case of the ordered rationals, their density everywhere and lack of endpoints. However, unlike in the back-and-forth proof where the new structural concept provides us with a general *technique* for showing isomorphism, the Fraïsse limit proof is essentially just the realization that $\langle \mathbb{Q}, < \rangle$ holds a special place in the family of linear orderings. As Fraïsse limits are unique for each class, we come to understand that any countable model of DLO must be isomorphic to the ordered rationals.

While the characterizing properties of $\langle \mathbb{Q}, < \rangle$ (*i.e.*, its countability and homogeneity) underpin both the back-and-forth and Fraïsse limit proofs, what is really significant and interesting is how these proofs exploit these properties, on one hand arguing that the characterizing properties of two dense linear orderings without endpoints facilitate the construction of a system of mappings between them, and on the other, recognizing these properties as features of a special limit structure among the linear orderings. Both proofs are driven by the introduction of new structural elements which either function as an instrument for investigating the theory of linear orderings (back-and-forth) or sharpen our understanding of how models of this theory are intimately related (in Fraïsse's terminology, some orderings are 'younger' or 'older' than others). This method of structural argumentation - proving mathematical facts by introducing and manipulating new structural concepts - is not unique to model theory either. In analysis, there are Cauchy sequences, metric spaces and coverings; in algebra, we decompose groups into cosets, examine eigenspaces of a matrix and so on. Mathematics is saturated with further examples.

In contrast to the structural proofs, the quantifier elimination argument proceeds at the syntactic level, tying results from the formal theory of DLO to actual orderings. The proof may seem a somewhat unusual one, deriving the structural result that any countable dense linear ordering without endpoints is isomorphic to the ordered rationals from the mere fact that the syntax needed to express the theory DLO need not include quantifiers. But model theorists Chang and Kiesler certainly feel that such syntactic methods *can* explain, writing in their discussion of preservation phenomena (where the syntactic form of a theory T can indicate whether T is preserved under submodels, unions of chains or homomorphisms): "It is a rather remarkable fact that these preservation phenomena can be explained just by the syntactic form of the axioms [of a theory]." (XXX) Nonetheless, I feel that characterizing the quantifier elimination proof as entirely syntactic would be somewhat misleading as several steps needed in proving DLO admits quantifier elimination, such as recognizing every open formula as DLO-equivalent to a finite disjunction of arrangements and showing $\exists x_n \Theta$ is equivalent (mod DLO) to Θ^*, appeal to structural knowledge of DLO. The quantifier elimination proof is rather an interplay between syntactic and structural reasoning (though I feel the heavy reliance on formal language still warrants the claim that this is a 'syntactic argument') which here operates to explain the ω-categoricity of DLO. Note that Henkin's completeness proof of first-order logic, proposed by Resnik and Kushner (1987: 149-50) as a counterexample to Steiner's model, also falls under this category.

The three model-theoretic proofs presented in this paper thus explain the same mathematical fact in the theory of linear orderings in diverse ways. In the proofs using back-and-forth equivalence and Fraïsse limits, the ω-categoricity of DLO is explained through structural arguments driven by the introduction of new relations and structural concepts. In the quantifier elimination proof, the same result is explained by analyzing the deductive consequences of the formal theory DLO. An account of mathematical explanation that simply tells us that these proofs explain, full stop, without flushing out the structural and syntactic methods catalyzing these explanations is unsatisfactory. *Even if* we accept that the distinctive feature of explanatory proofs is their dependence on the characterizing properties of mathematical objects, Steiner's model still fails us precisely because - though it tells us characterizing properties are all-important - it ignores the variety of methods (I have only indicated a few here) by which these properties become part of actual mathematical practice.

5 Final Remarks

If forced to give a concise answer to the question 'What is a mathematical explanation?', my answer would be this: a mathematical explanation is any piece of mathematical research, formal proof or otherwise, that raises understanding (as opposed to knowledge). Admittedly, I have replaced one vague notion, 'mathematical explanation', with another, 'mathematical understanding'. Yet, without getting into the details, recent work by Jeremy Avigad (forthcoming) has attempted to bring mathematical understanding down to earth, construing understanding as the possession of certain abilities, such as, in the case of group theory, being able to form a quotient group, list the finite groups of order less than 12 and so on. Whether such an account can ultimately be made to work (and I do think it is a good start), Avigad's model does indicate that, at least with regards to mathematical understanding, there is something to hold on to.

Accompanying the shift in focus from mathematical explanation to understanding under this proposal is a move away from *explanatory objects*, such as explanatory/non-explanatory proofs (Steiner) or unified/disjoint theories (Kitcher), towards the variety of mathematical methods that help us understand. We might still say that one proof explains while another does not but this is because one proof instantiates an *explanatory method*, or mode of reasoning that leads to understanding, while the other proof simply leads us to know that a certain result holds. In this case study from model theory, we saw examples of such explanatory methods in the introduction/manipulation of new structural concepts in the back-and-forth equivalence and Fraïsse limits proofs and the syntactic reasoning that showed that DLO admits quantifier elimination.

I am certainly not the only author to stress the heterogeneity of mathematical methods. Hafner and Mancosu (2005a: 222) write: "We maintain that mathematical explanations are heterogeneous...the variety of mathematical explanations cannot be easily reduced to a single model." In a recent talk at UC-Berkeley, the famous mathematician Dana Scott also expressed his doubts that the philosophy of mathematics, with such tidy accounts as Lakatos' 'Proofs and Refutations', has been able to accurately describe mathematical practice as in actuality "mathematicians are busy getting their hands dirty". Whether one agrees that mathematical explanations are best understood via mathematical understanding or not, I think a strong case has been made here that accounts of mathematical explanation, such as Steiner's, that proport to pin down the precise nature of mathematical explanation while ignoring the variety of diverse methods operative in the mathematical community are sweeping the dirt under the carpet, failing to provide an accurate picture of how real-life mathematicians explain mathematical facts.

6 References

Avigad, J. (forthcoming) *Understanding Proofs* [need complete reference]

Chang, C. & Keisler, H. J. (1973) *Model Theory (Studies in Logic and the Foundations of Mathematics)*, Elsevier Science B.V.: Netherlands.

Hafner, J. & Mancosu, P. (2005a) The Varieties of Mathematical Explanation, in P. Mancosu et al. (eds.), Visualization, Explanation and Reasoning Styles in Mathematics, Dordrecht: Springer.

Hafner, J. & Mancosu, P. (2005b) Explanation and Unification: A Case Study from Real Algebraic Geometry, talk presented at the conference 'Proof in Mathematics' Universite Lille III, May 24-28, 2005.

Hodges, W. (1997) *A Shorter Model Theory*, Cambridge University Press: Cambridge.

Kitcher, P. (1984) *The Nature of Mathematical Knowledge*, Oxford University Press: New York, Oxford.

Kitcher, P. (1989) Explanatory Unification and the Causal Structure of the World, in P. Kitcher & W. Salmon, (eds), *Scientific Explanation*, Vol. XIII of Minnesota Studies in the Philosophy of Science, University of Minnesota Press: Minneapolis.

Mancosu, P. (2001) Mathematical Explanation: Problems and Prospects, Topoi 20: 91-117.

Marker, D. (2002) *Model Theory: An Introduction*, Springer-Verlag: New York.

Resnik, M. & Kushner, D. (1987) Explanation, Independence and Realism in Mathematics, British Journal for the Philosophy of Science 38: 141-158.

Steiner, M (1978) Mathematical Explanation, Philosophical Studies 34: 135-151.