....

Minimization of Boolean complexity in human concept learning

Jacob Feldman

Department of Psychology, Center for Cognitive Science, Rutgers University, New Brunswick, New Jersey, 08903, USA

....

One of the unsolved problems in the field of human concept learning concerns the factors that determine the subjective difficulty of concepts: why are some concepts psychologically simple and easy to learn, while others seem difficult, complex or incoherent? This question was much studied in the 1960s[1] but was never answered, and more recent characterizations of concepts as prototypes rather than logical rules[2,3] leave it unsolved[4–6]. Here I investigate this question in the domain of Boolean concepts (categories defined by logical rules). A series of experiments measured the subjective difficulty of a wide range of logical varieties of concepts (41 mathematically distinct types in six families—a far wider range than has been tested previously). The data reveal a surprisingly simple empirical 'law': the subjective difficulty of a concept is directly proportional to its Boolean complexity (the length of the shortest logically equivalent propositional formula)—that is, to its logical incompressibility.

How human learners extract rules from patterns of data, and the relative subjective complexity of different kinds of rules, is central to investigations of human learning. When classifying novel stimuli, humans tend to seek simple rules rather than prototypes[7]. Subjective complexity has a central theoretical role in popular rule-plus-exception models of concept learning[8], and more generally in debates on the nature of learning[9,10], where it is often argued that examples failing to obey some 'simple' primary rule might be stored in an alternative manner, for example, verbatim. Earlier research focused on an extremely limited variety of logical forms, virtually exclusively comprising concepts defined with only two features. Most research considered just two types of logical rule, conjunction and disjunction, with one empirical result—that conjunctive ('and') concepts are learned more easily than disjunctive ('or') ones—dominating discussion. One early report[11] suggested a role for logical complexity, but this proposal was never followed up. Most theoretical accounts since then[1,8,12,13] have taken the subjective preference for conjunctive concepts as axiomatic; usually this is the only principle invoked pertaining to logical form. But no independently motivated explanation of it has ever emerged.
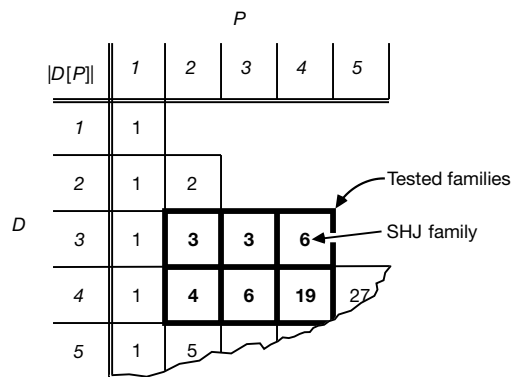
One study, that of Shepard, Hovland and Jenkins[14] investigated concepts with three features. Their stimulus classification and basic empirical result provides the motivation for my study. Shepard *et al.* considered Boolean concepts defined over three features with four positive and four negative examples. Such concepts fall into six logical types, designated I–VI (the SHJ types). All concepts with three features and four positive examples are isomorphic to one of the six types, but no example of one type is isomorphic to one of a different type, making this a complete classification. Illustrations of the six types can be found in Table 1 in the subtable labelled 3[4] (the notation is explained below). Shepard *et al.* found that the types differed in subjective difficulty in the order I < II < [III, IV, V] < VI, with types III, IV and V having approximately equal difficulty. This basic empirical pattern has been much discussed and occasionally replicated since[8,15], but while it has been modelled by simulations involving many free parameters, it has resisted any simple or elegant theoretical explanation.

When the SHJ types are considered from the perspective of mathematical logic, however, a simple explanation of the difficulty ordering emerges: the difficulty of the six types is precisely predicted by their Boolean complexity. The Boolean complexity of a propositional concept is the length of the shortest Boolean formula logically equivalent to the concept, usually expressed in terms of the number of literals (positive or negative variables)[16,17]. (For convenience, I write $a \wedge b$ as ab, a, $a \vee b$ as $a + b$, and $\neg a$ as a'.) For example, the concept $ab + ab'$ is equivalent to $a(b + b')$ and thus to a, and hence has Boolean complexity 1; whereas $ab + a'b'$ has no shorter equivalent, and hence has Boolean complexity 4. Like Kolmogorov complexity[18–21], Boolean complexity is an essentially universal measure of the intrinsic mathematical complexity or 'incompressibility' of the propositional concept[16]. Finding the shortest formula

Figure 1 The $D[P]$ hierarchy, showing the number of cases ($D[P]$) in each family. The hierarchy extends infinitely to all values of D, and all values of P less than or equal to 2^{D-1}. Families in the bold box are those tested in the experiments. Family 3[4] is the one originally considered by Shepard *et al.*[14].
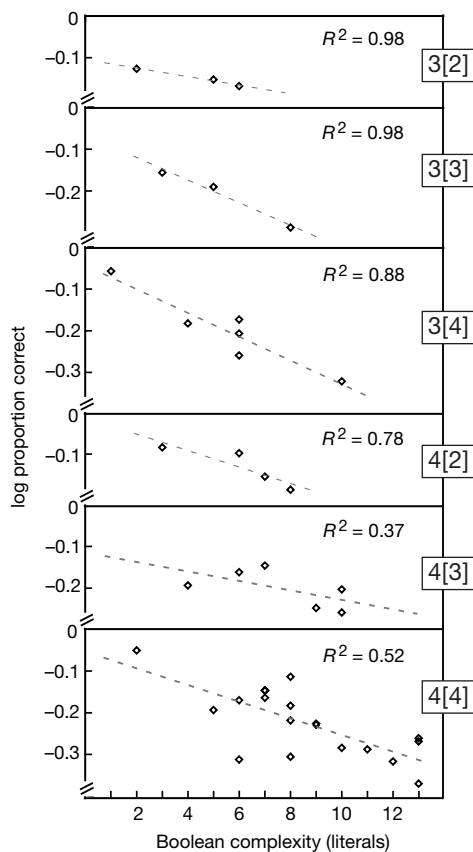

Figure 2 Results, showing log proportion correct as a function of Boolean complexity for each family (collapsing over parity). Full plots including all cases in both parities can be found in Supplementary Information.
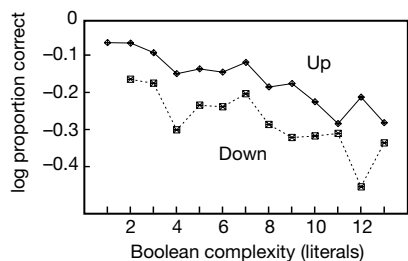
630 NATURE | VOL 407 | 5 OCTOBER 2000 | www.nature.com

Figure 3 Results by parity, collapsing over family. The 3[4] family is included with the Up cases. These results should be interpreted with caution because the different families may have different overall levels of difficulty (controlled in part by the amount of learning time, and only approximately equalized by the chosen levels).

equivalent to a given formula is computationally intractable[22], and in practice Boolean reduction can only be achieved through approximate computational techniques (such as factorization, as illustrated in the example).

Now consider an arbitrary Boolean concept defined by P positive examples over D binary features (for example, in the six SHJ types, $D = 3$ and $P = 4$). Any such concept is logically equivalent to the disjunction of its P constituent objects, each of which is a conjunction of D features, a form known as a disjunctive normal formula (DNF). Hence, each such concept is equivalent to a DNF with DP literals. For example, each of Shepard et al.'s six types is equivalent to a formula of length 12 ($= 3 \times 4$). The DNF is the completely 'uncompressed' form; it in effect lists verbatim all the objects that satisfy the concept. When each of the SHJ types is compressed as much as possible (again, using heuristic techniques), they have lengths 1, 4, 6, 6, 6, and 10, respectively (Table 1, section 3[4], gives

all the full DNFs and minimal formulae). These Boolean complexity values predict the order of empirical difficulty precisely. This exact correspondence has not previously been noted, though Shepard et al. speculated about it in their original paper[14], and the relation between Boolean complexity and human learning has never been comprehensively tested.

To investigate more thoroughly the role of Boolean complexity in learning, the next natural step is to extend the range of propositional concepts beyond Shepard et al.'s original six. Those six cases exhaust all possibilities with $D = 3$ and $P = 4$; an obvious extension is to consider different values of D, P, or both. Such changes lead to completely different classifications. For example, with $D = 3$ and $P = 3$ there are three cases, each necessarily distinct from any of the six used by Shepard et al.

Varying both D and P defines a parametric system of families, each analogous to, but different from, the SHJ family. I use the notation $D[P]$ to indicate the family of concept types involving D features and P positive examples, and denote the cases within that family by $I_{D[P]}$, $II_{D[P]}$ and so on. Under this notation, the SHJ family is 3[4], and its six cases are $I_{3[4]}$, $II_{3[4]}, \ldots, VI_{3[4]}$. Figure 1 gives the number of cases $|D[P]|$ for each family from $D = 1$ to 4 and for $P = 1$ to 2^{D-1}. Values of P larger than 2^{D-1} duplicate smaller values; for example, a concept with three positives and five negatives is simply the converse of one with five positives and three negatives, and hence, has essentially the same form and the same complexity.

The $D[P]$ hierarchy is completely comprehensive, in that it encompasses the entire universe of propositional formulae. *All* Boolean concepts fall into a family that appears somewhere in (the infinite extension of) Fig. 1, and *all* Boolean concepts are logically equivalent to one somewhere on the (infinite extension of) Table 1. Hence, the $D[P]$ complexity hierarchy is the natural testbed in which to evaluate any psychological theory that makes reference

Table 1 Details of two concept families tested in the experiments

Family*	Case	DNF	Minimal formula	Complexity	Illustration
3[3]	$I_{3[3]}$	$a'b'c' + a'b'c + a'bc'$	$a'(bc)'$	3	
	$II_{3[3]}$	$a'b'c' + a'b'c + abc'$	$a'b' + abc'$	5	
	$III_{3[3]}$	$a'b'c' + a'bc + ab'c$	$a'(b'c' + bc) + ab'c$	8	
3[4]	$I_{3[4]}$	$a'b'c' + a'b'c + a'bc' + a'bc$	a'	1	
	$II_{3[4]}$	$a'b'c' + a'b'c + abc' + abc$	$ab + a'b'$	4	
	$III_{3[4]}$	$a'b'c' + a'b'c + a'bc' + ab'c$	$a'(bc)' + ab'c$	6	
	$IV_{3[4]}$	$a'b'c' + a'b'c + a'bc' + ab'c'$	$a'(bc)' + ab'c'$	6	
	$V_{3[4]}$	$a'b'c' + a'b'c + a'bc' + abc$	$a'(bc)' + abc$	6	
	$VI_{3[4]}$	$a'b'c' + a'bc + ab'c + abc'$	$a(b'c + bc') + a'(b'c' + bc)$	10	

For each type the Roman-numeral case label, disjunctive normal formula, minimal formula, Boolean complexity, and a schematic illustration of the concept as a set of vertices in Boolean D-space are shown. Minimal formulae and complexity values are derived using heuristic minimization techniques, such as factoring (for details see Supplementary Information). All concepts with D features and P positives can be created by taking one of these cases and permuting and/or inverting the axes in Boolean D-space — equivalent to permuting the features and the parity of each feature. Ordering of cases within each family is arbitrary.
* Family 3[4] is the original family from Shepard et al.[14]. For details of other families see Suplementary Information.

to the logical structure of concepts.

The experiments below focus on six families: 3[2], 3[3], 3[4] (the SHJ family), 4[2], 4[3] and 4[4]; that is, all concepts with three or four features and between two and four positive examples, comprising 41 cases in the six families. Table 1 gives details of families 3[3] and 3[4]. Each concept (except in 3[4], where positive and negative sets are the same size) was tested in two versions: an Up version, in which the smaller set of objects was labelled 'positive', and a Down version, in which the larger set was labelled 'positive'. This factor, which is orthogonal to Boolean complexity (because every concept ϕ and its complement ϕ' have the same complexity) will be referred to as parity.

Results are plotted in Fig. 2 (separated by family) and Fig. 3 (separated by parity). The main trend is that performance decreases monotonically with increasing Boolean complexity, with an approximately constant advantage for Up versus Down cases. Together the complexity and parity effects explain about half of the variance ($R^2 = 0.5017$, $F(2, 73) = 36.76$, $P < 0.0001$). The parity effect, which is clearly visible in all families, has antecedents in the literature[12,23–25]. Again, this factor is orthogonal to complexity and thus cannot be explained by complexity minimization.

My main conclusion is that subjective conceptual difficulty is well predicted by Boolean complexity. For each concept, learning is successful to the degree that the concept can be faithfully compressed[26,27]. However, complexity and parity alone do not provide a complete account; there is substantial residual variance in the data (especially in higher-dimensional cases, and in Down cases in all dimensions) suggesting the influence of some unknown processing factors and strategies, which require future study.

The significance of these results is best viewed from an historical perspective. Concept learning research in the 1960s was largely preoccupied by the dichotomy between conjunctive and disjunctive concepts. In retrospect, much of the interest in this comparison derived from the fact that it seemed to imply a divergence between logical complexity, in which the two types are manifestly equal, and psychological complexity, in which the two types differ markedly. The contrast suggested a predominant role of extra-logical, and hence perhaps more subjective and mysterious, factors.

But in the light of the current study, this conclusion seems to have been premature. In the $D[P]$ classification, conjunction ab and disjunction $a + b$ represent the same case, $\text{I}_{2[1]}$; conjunction is the Up version (one positive) and disjunction is the Down version (three positives). That is, conjunction and disjunction are isomorphic, in that each divides Boolean 2-space into one vertex versus the other three; in conjunction the one is labelled positive and the rest negative, in disjunction the reverse. (Conditional $a \rightarrow b$ is also of the same type.) All other propositional concepts studied in the literature (again, apart from ref. 14) are bivariate concepts falling into the 2[2] family, which has only two cases, each appearing in two complementary versions: affirmation a and negation a' (both $\text{I}_{2[2]}$); and biconditional $a \leftrightarrow b$ (that is, $a'b' + ab$) and exclusive disjunction $ab' + a'b$ (both $\text{II}_{2[2]}$). Conjunction/disjunction, affirmation/negation and biconditional/xor have Boolean complexities 2, 1 and 4, respectively, approximately agreeing with their empirical difficulty ordering[1] once parity is taken into account. (This remark simply recapitulates Neisser and Weene's paper[11] in modern terminology; perhaps their paper would have had more influence if parity had been recognized as an orthogonal factor to complexity, thus improving agreement with the data.) In this light the subjective advantage of conjunction over disjunction admits a completely different explanation: the preference for Up over Down parity. If this account is correct, it places a very different interpretation on the main research result of this literature. In effect the traditional argument that logical complexity does not predict subjective difficulty was based on a few relatively simple cases that sit at the tip of a very large iceberg. The rest of the iceberg—the full $D[P]$ hierarchy—tells a very different story.

In a sense, this final conclusion may seem negative: human conceptual difficulty reflects intrinsic mathematical complexity after all, rather than some idiosyncratic and uniquely human bias. The positive corollary though is certainly more fundamental: subjective conceptual complexity can be numerically predicted and perhaps explained. The next step is to reconcile this finding with contemporary theory on concepts involving prototypes or other more realistic types of category. □

Methods

1. Bourne, L. E. Knowing and using concepts. *Psychol. Rev.* **77**, 546–556 (1970).
2. Posner, M. I. & Keele, S. W. On the genesis of abstract ideas. *J. Exp. Psychol.* **77**, 353–363 (1968).
3. Rosch, E. H. Natural categories. *Cogn. Psychol.* **4**, 328–350 (1973).
4. Osherson, D. N. & Smith, E. E. On the adequacy of prototype theory as a theory of concepts. *Cognition* **9**, 35–58 (1981).
5. Medin, D. L. & Smith, E. E. Concepts and concept formation. *Annu. Rev. Psychol.* **35**, 113–138 (1984).
6. Komatsu, L. K. Recent views of conceptual structure. *Psychol. Bull.* **112**, 500–526 (1992).
7. Medin, D. L., Wattenmaker, W. D. & Hampson, S. E. Family resemblance, conceptual cohesiveness, and category construction. *Cogn. Psychol.* **19**, 242–279 (1987).
8. Nosofsky, R. M., Palmeri, T. J. & McKinley, S. C. Rule-plus-exception model of classification learning. *Psychol. Rev.* **101**, 53–79 (1994).
9. Pinker, S. Rules of language. *Science* **253**, 530–535 (1991).
10. Smith, E. E., Patalano, A. L. & Jonides, J. Alternative strategies of categorization. *Cognition* **65**, 167–196 (1998).
11. Neisser, U. & Weene, P. Hierarchies in concept attainment. *J. Exp. Psychol.* **64**, 640–645 (1962).
12. Bourne, L. E., Ekstrand, B. R. & Montgomery, B. Concept learning as a function of the conceptual rule and the availability of positive and negative instances. *J. Exp. Psychol.* **82**, 538–544 (1969).
13. Medin, D. L., Wattenmaker, W. D. & Michalski, R. S. Constraints and preferences in inductive learning: an experimental study of human and machine performance. *Cogn. Sci.* **11**, 299–339 (1987).
14. Shepard, R., Hovland, C. L. & Jenkins, H. M. Learning and memorization of classifications. *Psychol. Monogr. Gen. Appl.* **75**, 1–42 (1961).
15. Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C. & Glauthier, P. Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland and Jenkins (1961). *Mem. Cogn.* **22**, 352–369 (1994).
16. Wegener, I. *The Complexity of Boolean Functions* (Wiley, Chichester, 1987).
17. Givone, D. D. *Introduction to Switching Circuit Theory* (McGraw Hill, New York, 1970).
18. Chaitin, G. J. On the length of programs for computing finite binary sequences. *J. Assoc. Comput. Machin.* **13**, 547–569 (1966).
19. Kolmogorov, A. N. Three approaches to the quantitative definition of information. *Problems Inf. Transm.* **1**, 1–7 (1965).
20. Solomonoff, R. J. A formal theory of inductive inference. Part I. *Inf. Control.* **7**, 1–22 (1964).
21. Li, M. & Vitányi, P. *An Introduction to Kolmogorov Complexity and its Applications* (Springer, New York, 1997).
22. Garey, M. R. & Johnson, D. S. *Computers and Intractability: A Guide to the Theory of NP-completeness* (Freeman, New York, 1979).
23. Hovland, C. I. & Weiss, W. Transmission of information concerning concepts through positive and negative instances. *J. Exp. Psychol.* **45**, 175–182 (1953).
24. Haygood, R. C. & Devine, J. V. Effects of composition of the positive category on concept learning. *J. Exp. Psychol.* **74**, 230–235 (1967).
25. Nahinsky, I. D. & Slaymaker, F. L. Use of negative instances in conjunctive concept identification. *J. Exp. Psychol.* **84**, 64–68 (1970).
26. Rissanen, J. Modeling by shortest data description. *Automatica* **14**, 465–471 (1978).

27. Chater, N. Reconciling simplicity and likelihood principles in perceptual organization. *Psychol. Rev.* **103**, 566–581 (1996).

....

Skin abnormalities generated by temporally controlled RXRα mutations in mouse epidermis

Mei Li*†, **Arup Kumar Indra***†, **Xavier Warot***, **Jacques Brocard***, **Nadia Messaddeq***, **Shigeaki Kato**‡§, **Daniel Metzger***
& Pierre Chambon*

* *Institut de Génétique et de Biologie Moléculaire et Cellulaire,*
CNRS/INSERM/ULP, Collège de France, BP 163, 67404 Illkirch Cedex, France
‡ *Institute of Molecular & Cellular Biosciences, University of Tokyo, Yayoi,*
Bunkyo-ku, Tokyo 113, Japan
§ *CREST, Japan Science and Technology, Kawaguchi, Saitama 332, Japan*
† *These authors contributed equally to this work*

....

Nuclear receptors for retinoids (RARs) and vitamin D (VDR), and for some other ligands (TRs, PPARs and LXRs), may be critical in the development and homeostasis of mammalian epidermis[1–8]. It is believed that these receptors form heterodimers with retinoid X receptors (RXRs) to act as transcriptional regulators[9,10]. However, most genetic approaches aimed at establishing their physiological functions in the skin have been inconclusive owing either to pleiotropic effects and redundancies between receptor isotypes in gene knockouts, or to equivocal interpretation of dominant-negative mutant studies in transgenic mice[1,13–15]. Moreover, knockout of RXRα, the main skin RXR isotype, is lethal *in utero* before skin formation[11,12,16,17]. Here we have resolved these problems by developing an efficient technique to create spatio-temporally controlled somatic mutations in the mouse. We used tamoxifen-inducible Cre–ER[T] recombinases[18,19] to ablate RXRα selectively in adult mouse keratinocytes. We show that RXRα has key roles in hair cycling, probably through RXR/VDR heterodimers, and in epidermal keratinocyte proliferation and differentiation.

To ablate RXRα in epidermis, we engineered mice carrying LoxP-site-containing (floxed) RXRα[L2] alleles (Fig. 1a) and used the K5–Cre–ER[T] transgenic line in which tamoxifen (Tam) efficiently induces Cre-mediated recombination in basal layer keratinocytes[19]. K5–Cre–ER[T(tg/tg)]/RXRα[L2/L2] mice mated with RXRα[+/−] (Fig. 1a; ref. 16) or RXRα[L2/+] mice yielded 'pro-mutant' mice hemizygous (tg/0) for K5–Cre–ER[T] and carrying either one RXRα[L2] and one RXRα null (−) allele (K5–Cre–ER[T(tg/0)]/RXRα[L2/−] genotype) or two L2 alleles (K5–Cre–ER[T(tg/0)]/RXRα[L2/L2] genotype). At 14 weeks old, the pro-mutant mice were treated with Tam (5 days, 1 mg per day), and then retreated 2, 4 and 6 weeks later. Six weeks after the first Tam treatment (AFT), 80% of RXRα[L2] alleles were converted into RXRα[L−] alleles in the epidermis of mice carrying one or two floxed alleles (Fig. 1b). By 12 weeks AFT, almost all RXRα[L2] alleles had been

converted (Fig. 1b). As expected[19], no RXRα disruption occurred in vehicle (oil)-treated mice (data not shown) and Cre-mediated excision of RXRα exon 4 was restricted to epidermis and some epithelia in which the K5 promoter is also active (for example, tongue, salivary gland, oesophagus; Fig. 1c).

Interestingly, hair loss (alopecia) was observed 6–7 weeks AFT in the ventral region of pro-mutant mice, but not in oil-treated pro-mutant mice or in Tam-treated K5–Cre–ER[T(tg/0)]/RXRα[L2/+] 'control' littermates (data not shown). At 12–16 weeks AFT, large regions of ventral skin and smaller regions of dorsal skin were hairless (Fig. 2a, b; and data not shown). Cysts became visible under the skin surface and these enlarged and spread all over the body with time (Fig. 2c; and data not shown). With increasing age (> 20 weeks AFT), minor focal lesions appeared on hairless dorsal skin, on chins and behind ears (Fig. 2d; and data not shown). These were not caused by fights and were formed of crusts on top of hyperproliferative epidermis and inflammatory dermis (see below).
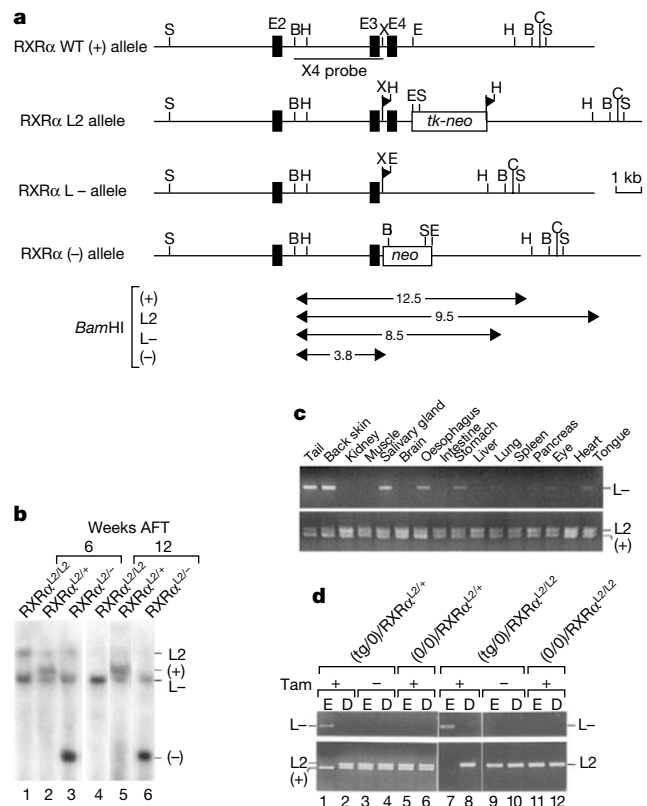

Figure 1 Tamoxifen-induced RXRα null mutation in adult mouse epidermis mediated by Cre–ER[T]. **a**, Diagram of the wild-type RXRα genomic locus (+), the floxed RXRα L2 allele, the RXRα L– allele obtained after Cre-mediated excision of exon 4 (encoding the DNA-binding domain), and the RXRα null allele (−)[16]. Black boxes indicate exons (E2–E4). Restriction enzyme sites and probe X4 location are indicated. *Bam*HI fragments are in kilobases (kb). B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII, S, *Spe*I; X, *Xba*I. Arrowheads in L2 and L– alleles indicate LoxP sites. **b**, Tamoxifen (Tam)-induced generation of K5–Cre–ER[T]-mediated RXRα[L–] alleles illustrated by Southern blot analysis of epidermal DNA isolated 6 (lanes 1–3) and 12 (lanes 4–6) weeks after the first Tam (1 mg) injection series (AFT). All mice were K5–Cre–ER[T(tg/0)] and the RXRα genotypes are indicated. *Bam*HI-digested DNA fragments corresponding to RXRα (+), L2, L– and (−) alleles are displayed. **c**, Tissue-specificity of Cre-ER[T]-mediated RXRα disruption. WT (+), L2 and L– alleles were identified by PCR on DNA extracted from various organs of K5–Cre–ER[T(tg/0)]/RXRα[L2/+] mice, 12 weeks AFT. **d**, Tamoxifen-induced generation of RXRα null alleles in adult mouse epidermis using K14–Cre–ER[T2(tg/0)] or K14–Cre–ER[T2(0/0)] mice (designated (tg/0) and (0/0), respectively). PCR analysis of genomic DNA from epidermis (E) and dermis (D), isolated two weeks after injection of either Tam (0.1 mg) (+) or vehicle (−). Mouse genotypes are indicated and PCR fragments corresponding to RXRα (+), L2 and L– alleles are displayed.