

Towards a quantitative approach for objectively measuring the similarity of marks

by David Barnett

Introduction

Central to many intellectual property disputes is an assessment of the degree of similarity of two contested marks. This follows from the framework set out in (for example) the UK Trade Marks Act 1994¹, whereby even a non-identical mark may be considered non-registrable or infringing if it creates a likelihood of confusion (Sections 5(2b) and 10(2b)) with an earlier mark².

A key point to note is that decisions regarding legal similarity are fundamentally subjective, involving a range of relevant tests which include consideration of the perception of the relevant consumer, and recognition of the existence of degrees of similarity within a spectrum (from high to low).

Whilst trademark comparisons are likely always to retain a degree of subjectivity, there are some areas where objective quantitative formulations can be constructed. A more objective framework could have a number of advantages, including the potential to quantitatively measure the difference between marks. It would be necessary to explicitly incorporate the relevant metrics into comparison tests, but this would offer the potential to define thresholds up to which IP protection could apply, and could provide the basis for new case law to be applied to future analogous disputes, offering the potential for greater legal consistency.

In this article, I consider the cases of colour- and word marks, and outline some potential methodologies for quantifying the degree of similarity of marks.

It is worth noting that a number of trademark search tools already offer automated comparison screening, and generally require the review of an experienced practitioner to prioritise the results, remove irrelevant false positives, and draw the results into a legal analysis. The comparison frameworks presented in this article doubtless have some parallels with the algorithms used by these types of providers (and are compared against one such example in a forthcoming article in this series). All of these types of quantitative comparison approaches are likely to continue to need to be accompanied by manual review, and the formulations presented in this article are suggested merely as tools to be incorporated into existing approaches and doctrines, rather than being intended to *replace* (on a wholesale basis) the current nuanced and multi-faceted approach to infringement employed by courts and tribunals.

¹ <https://assets.publishing.service.gov.uk/media/63f8963e8fa8f527f110a2e6/Consolidated-Trade-Marks-Act-1994-February23.pdf>

² Or if the mark “take[s] unfair advantage of, or [is] detrimental to the distinctive character or ... repute” (Sections 5(3) and 10(3)) of, the earlier mark

Part 1: Colour, similarity, and intellectual property

Introduction: colour as a protected characteristic

The history and case-law surrounding trademarks as an indicator of product origin is extremely well established. Trademarks most usually pertain to brand names (word marks) or figurative elements such as logos. There are, however, a number of other characteristics which can have powerful brand associations, including product 'look-and-feel' (trade dress), sound marks and colours.

The legal definition of a trademark was broadened by the World Trade Organization Agreement on Trade-Related Aspects of Intellectual Property Rights, to cover "any sign ... capable of distinguishing ... goods or services"³, and many specific colours have been registered as trademarks by corporations (either as colour marks *per se*, or as colours included as components of a more complex mark featuring additional elements) (Figure 1) for their particular product classes. The most popular colour groups to be registered as trademarks are shades of blue (18% of registrations), red or pink (18%), yellow or gold (15%), and green (14%). Registration is possible if the colour serves as an indication of source, if it is not purely decorative or functional, and if proof of 'secondary meaning' can be provided (strictly, these definitions relate to US legal tests, which are relevant to various of the brands discussed below). In essence, this means that the public has come to associate the colour with the associated brand, which can be demonstrated through the use of extensive advertising featuring the colour and/or through consumer surveys⁴.



Figure 1: Examples of colours registered as trademarks by brand owners (either as full marks or as components of a more complex mark) (source: Z. Crockett / The Hustle⁵)

³ https://en.wikipedia.org/wiki/Colour_trade_mark

⁴ <https://ipwatchdog.com/2018/07/14/can-you-trademark-a-color/id=99237/>

⁵ <https://thehustle.co/can-a-corporation-trademark-a-color>

Some brands have been particularly enthusiastic in attempting to protect *ranges* of shades, with Deutsche Telekom claiming rights over a selection of variants of magenta for the T-Mobile brand (actually the subject of a recent case in the Benelux⁶), for example⁷.

One lesser-known illustration of the strength and appeal of distinctive colours in branding is the case of the Veuve Clicquot champagne brand. Veuve Clicquot utilises a striking yellow-orange colour for its label and packaging, and originally applied to register the colour as a protected mark in 1998 (Figure 2).

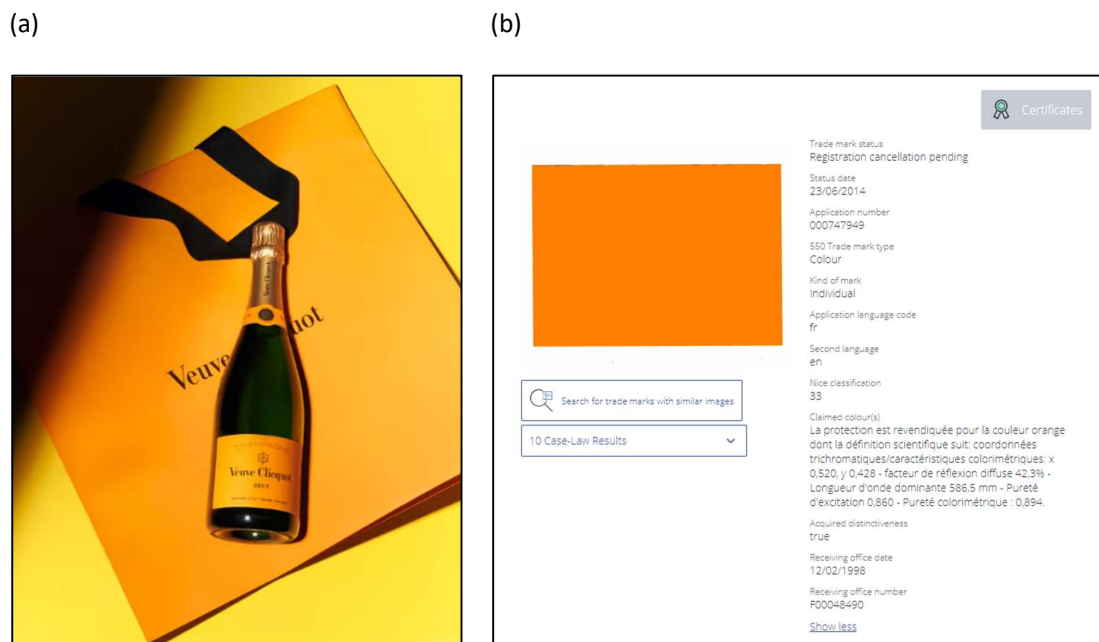


Figure 2: (a) The Veuve Clicquot product, as displayed on their official website (veuvecliquot.com/en-gb/home); (b) the registered colour mark⁸

Significantly, the colour was specified according to a scientific definition⁹, and was also erroneously stated to be a figurative registration (meaning that technically the protected mark was actually (exactly) an orange rectangle). As a result, the application was initially refused, before the refusal was annulled in 2002, with the registration considered to be that of a colour mark (now assigned to brand owner MCHS, a subsidiary of LVMH) (though still retaining an error whereby the dominant wavelength is given as 586.5 *mm* (10^{-3} m), rather than *nm* (10^{-9} m)). Subsequent cases have (unsuccessfully) attempted to invalidate the mark, including most recently (in 2018) by Lidl – renowned for their appetite for producing ‘lookalike’ products – followed by a 2024 attempt by Lidl to annul the refusal. Part of their case related to the complexity of the mark definition, which is rather different from the more familiar, accessible (and modern) approach of defining colours using Pantone, RGB (red-green-blue) or CMYK (cyan-magenta-yellow-black) codes.

⁶ <https://trademarkblog.kluweriplaw.com/2024/09/24/acquired-distinctiveness-for-non-traditional-trademarks-in-the-benelux-show-your-true-colours/>

⁷ <https://www.colourstudies.com/blog/2022/4/17/trademarking-colours>

⁸ <https://www.tmdn.org/tmview/#/tmview/detail/EM50000000747949>

⁹ Trichromatic coordinates / colour characteristics: x 0.520, y 0.428; diffuse reflectance 42.3%; dominant wavelength 586.5 nm [*sic*], excitation purity 0.860; colorimetric purity: 0.894

In another case, Cadbury’s attempt to register the purple shade Pantone 2685C (RGB [50,0,110]¹⁰; see explanation below) as a UK trademark for use on its chocolate packaging led to an opposition by Nestlé, with consideration given to the non-specificity regarding how the colour was to be applied, before the two parties ultimately reached a settlement. However, the case provides a clear precedent regarding the possibility of registering particular colours as trademarks¹¹.

As discussed above, a key element of these types of case is the requirement for the brand owner to be able to demonstrate ‘acquired distinctiveness’. One component of this objective is education of the public that the colour can function as a mark – i.e. a distinctive characteristic – in its own right. Initiatives along these lines have recently been employed by both Coca Cola and Mattel (for the Barbie brand), through the use of marketing campaigns incorporating minimalist brand references, where brand colours feature first and foremost¹² (Figure 3).

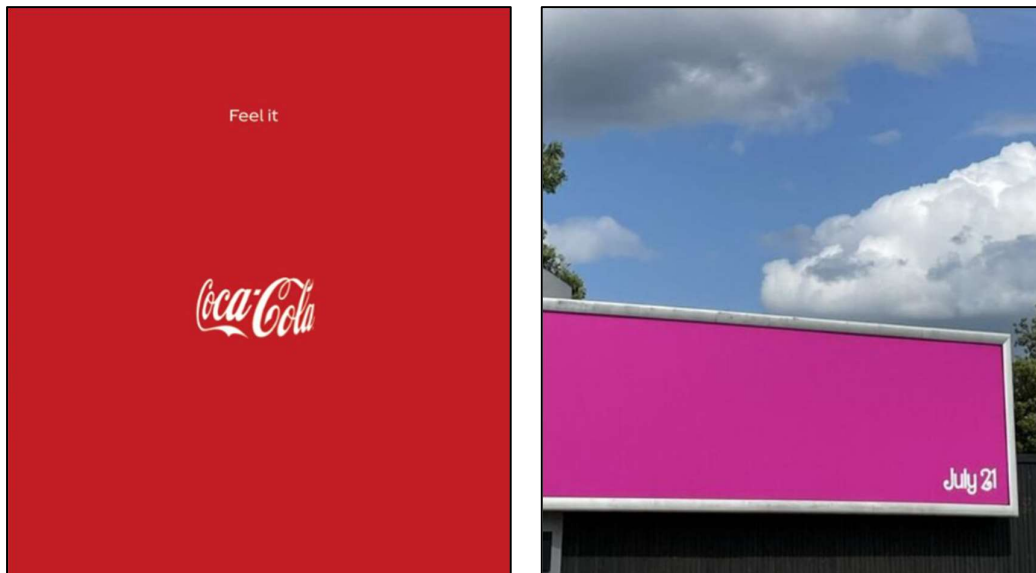


Figure 3: Marketing campaigns incorporating prominent display of brand colours for Coca Cola (left) and Barbie (right)

All these cases raise the question as to the effectiveness of the protection afforded by a registered colour mark. Is there anything to stop (for example) a third party using a shade which differs from a protected colour by (say) one RGB-point? Is there a threshold as to how close a colour needs to be to another, in order to be covered by the umbrella of protection (beyond the vague description that the similarity should be such that the difference between the shades is ‘barely noticeable’¹³)? For example, should the colour schemes for Stihl (a combination of orange RAL 2010 (RGB [208,93,40]¹⁴)

¹⁰ <https://www.namebadgesinternational.us/faqs/pantone-to-rgb/>

¹¹ <https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2022/1671.html> (referenced at: <https://www.pinsentmasons.com/out-law/analysis/cadbury-ruling-guide-registering-colour-trade-marks>)

¹² <https://www.linkedin.com/posts/rebecca-newman-267a7756-thinking-of-using-a-non-standard-trade-mark-activity-7216747278231883779-6SH3>

¹³ ‘Infringement of Colour Trademarks’, *GRUR International*, Vol. 70, Iss. 7 (2021), pp. 676–680 (<https://doi.org/10.1093/grurint/ikab061>) (available at: <https://academic.oup.com/grurint/article-abstract/70/7/676/6303754?>)

¹⁴ <https://rgb.to/ral/2010>

and grey RAL 7035 (RGB [197,199,196]¹⁵)^{16,17}) and Chinese manufacturer Emas (Figure 4) be allowed to co-exist without being deemed to create confusion¹⁸? The short response is that – currently – there is insufficient case law to provide a definitive answer, and that colour-mark protection is not enormously robust.

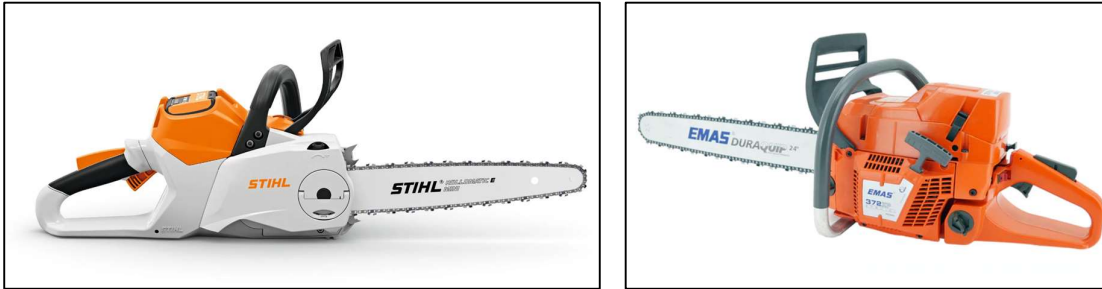


Figure 4: Stihl and Emas power-tool products, both of which use an orange-and-grey colour scheme

The situation is further complicated by the variability which may exist across the *same* product in a range of contexts (e.g. where different product imagery may be displayed on distinct websites, or even just when viewed on different devices) For example, Figure 5 shows four examples of listings for Cadbury’s iconic Dairy Milk product (nominally expected all to be Pantone 2685C), and the actual RGB values are different in each case.

¹⁵ <https://rgb.to/ral/7035>

¹⁶ <https://www.iam-media.com/article/stihl-successfully-invalidates-infringers-colour-combination-mark>

¹⁷

<https://curia.europa.eu/juris/document/document.jsf?text=&docid=239252&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=1772718> (referenced at:

<https://www.fieldfisher.com/en/services/intellectual-property/intellectual-property-blog/cutting-through-the-issues-colour-trade-marks-and> and <https://www.wiggin.co.uk/insight/general-court-annuls-board-of-appeal-decision-that-colour-combination-mark-was-not-sufficiently-clear-and-precise-to-indicate-origin/>)

¹⁸ <https://asiaiplaw.com/index.php/article/defending-stihls-orange-and-grey-colour-combination>



Figure 5: Examples of Cadbury’s Dairy Milk product images with differing RGB values (see explanation below) for the purple packaging (according to a colour analysis tool, on a central area of the packaging): (a) [74,25,114]; (b) [72,24,108]; (c) [50,40,130]; (d) [76,40,136]

However, the question of colour protection is an important one to be able to answer. Previous work, utilising consumer psychology studies, has shown that colour is one of the primary characteristics (together with packaging shape and style) used by consumers to identify products – with a greater importance than brand name. Colour increases brand recognition by 80%, and accounts for between 62% and 90% of a consumer’s initial judgement of a product¹⁹. These observations are the reason why the problem of lookalike products is such a concerning issue²⁰. Lookalikes can most effectively be addressed through the registration of packaging as a trademark and a subsequent unfair advantage claim, but the application of colour marks could also be part of the picture.

Removing the subjectivity

In many intellectual property disputes, a central component is the assessment of the degree of similarity (which, in reality, exists as a ‘spectrum’) between two marks. Whilst this is frequently a subjective determination, colour marks are somewhat different, in that colours can essentially be exactly defined, so that a quantitative measure of difference can be formulated. This being the case, colour marks should be able to lend themselves – through the development of an appropriate landscape of case law – to the formulation of a legal framework whereby similarity can be objectively measured, and thresholds up to which protection may apply, can be defined.

(As referenced above,) one simple model for defining colours (as presented in digital format) is the RGB framework, expressing the red, green and blue components (respectively) of any given colour as

¹⁹ <https://www.semanticscholar.org/paper/The-Psychology-of-Colour-Influences-Consumers%E2%80%99-%E2%80%93-A-Kumar/f7c3b2a780a7a3bf907ef807085b86a63f0d8d0a?p2df>

²⁰ <https://www.iamstobbs.com/the-psychology-of-lookalikes>

a number from 0 to 255, and thereby formulating a full ('3D') colour 'space' from [0,0,0] (black) to [255,255,255] (white), or 16,777,216 (i.e. 256^3) colours in total^{21,22}. Even this framework does not account for all possibilities, as there will be an infinite number of intermediate shades between any two adjacent RGB colours (if defined using integer numbers), and other characteristics (such as metallicness, reflectivity – i.e. the difference between 'gloss' and 'matt(e)' shades – or fluorescent properties) which are not fully accounted for – but the *idea* is the key point.

The model is such that any given colour is represented by a point in the colour space, as shown in Figure 6.

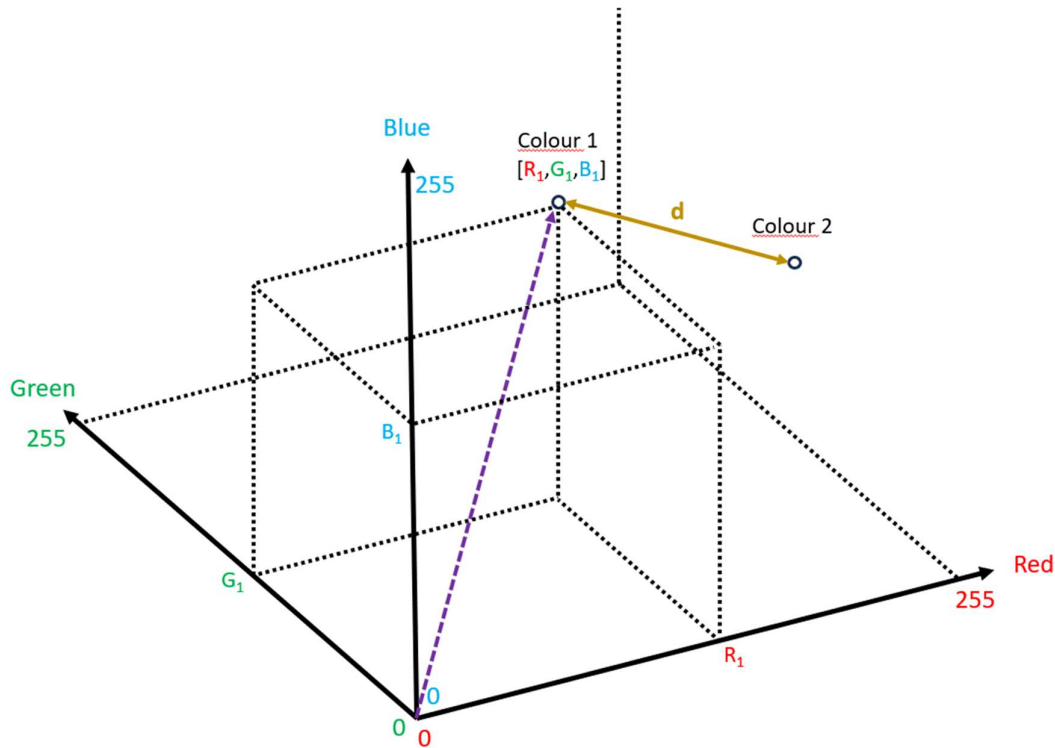


Figure 6: 3D representation of a colour ('Colour 1'; $[R_1, G_1, B_1]$) in RGB space

²¹ Equivalently, RGB values can be expressed in 'Hex' (hexadecimal, or base-16) format, where each component (R, G, and B) is expressed as a two-digit hexadecimal number, with each digit in the range from 0 to F (= 15). 255 is would therefore be expressed as FF (i.e. $15 \times 16^1 + 15 \times 16^0$), and [255,255,255] written as #FFFFFF

²² Interestingly, a crowdsourced initiative to assign a name to every colour in this space is currently underway at colornames.org. [I've suggested that [188/125/97] should be 'David Barnett's Face' – please lend your support and vote 😊]

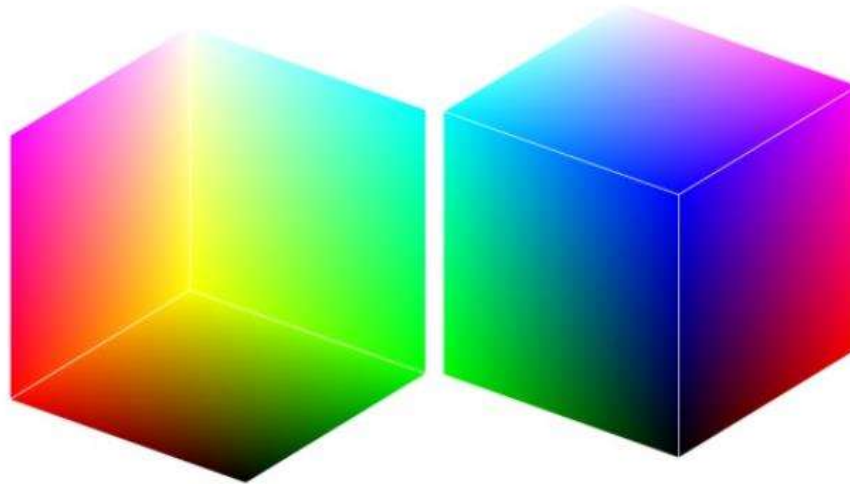


Figure 7: Visualisation²³ of the individual colours represented within RGB space (right-hand view of cube in approximately the same orientation as Figure 6)

This framework means that the similarity between two colours (i.e. the geometric distance between them in colour space – ‘*d*’ in Figure 6 – with a smaller value of *d* denoting colours which are more similar) can be exactly defined. Mathematically (according to Pythagoras’ theorem):

$$d = \sqrt{ (R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2 }$$

Consequently, the distance (*d*) between any two colours will be somewhere in the range of 0 to 442 (= $\sqrt{3 \times 255^2}$); the distance between black and white).

As an illustration, Table 1 shows the distances between each of the pairs of purple colours in the product images shown in Figure 5.

		Colour 1			
		[74,25,114]	[72,24,108]	[50,40,130]	[76,40,136]
Colour 2	[74,25,114]	-	6	33	27
	[72,24,108]		-	35	32
	[50,40,130]			-	27
	[76,40,136]				-

Table 1: Distances (*d*) (in RGB units) between each of the pairs of colours in the product images shown in Figure 5

This degree of variability, even across differing representations of the same product, is quantitatively relatively large – consistent with the marked visual differences. Since colour marks have the potential to be so oppressive (in terms of the limitations they may impose on third-party marks), it is probably reasonable that businesses should not be able to rely on (such significant) variations in colour across their own product, in order to define the threshold up to which protection should apply. However, it is reasonable that some limited range *should* be covered under the umbrella of

²³ <https://masacd.wordpress.com/svg-watercolor-cube/>

protection provided by a registered colour mark, due to variations in printing and digital display technologies. One implication of this approach would be that it would set an upper limit on the total number of colours within RGB space which could be protected²⁴ (much lower than the total ‘universe’ of 16.8 million colours).

As an example illustration of the extent of colour variation which exists as a function of distance (d), Figure 8 shows a series of visualisations of the colour space surrounding the point at RGB [50,0,110] (i.e. Cadbury’s Pantone 2685C). The circles in the figure show progressively increasing values of d , in steps of 10 units, up to a maximum of 50 (e.g. a protected colour-mark ‘bubble’ covering up to $d = 10$ would encompass the colour variations contained within the *innermost* circle).



Figure 8: Slices through RGB colour space surrounding the point at RGB [50,0,110]:

- (a) (*left*) slice perpendicular to the red axis, with green and blue increasing to the right and top, respectively
- (b) (*middle*) slice perpendicular to the green axis, with red and blue increasing to the right and top, respectively
- (c) (*right*) slice perpendicular to the blue axis, with green and red increasing to the right and top, respectively

Areas where any of the (R,G,B) parameters are less than zero or greater than 255 (i.e. those falling outside the colour space) are shown in black. Circles show progressively increasing values of d , in steps of 10 units, up to a maximum value of 50 units

It is certainly also possible to formulate modifications to the above approach, such as the use of mathematical weightings to account for the way in which colours are *perceived* through human vision²⁵, or the use of constructs such as the ‘normalised inner product’²⁶ (essentially, a measure of (just) the differences in ‘direction’ of each colour from the origin point (i.e. [0,0,0] or black) – as shown by the dashed purple vector arrow in Figure 6). However, arguably, the exact formulation is unimportant; the key point is that, provided a consistent methodology is used, objective measurements can be applied to the comparison assessment.

In may also be the case that similar approaches can be formulated for other types of marks whose assessment has traditionally been seen as much more subjective. Certainly, we are already seeing

²⁴ This upper limit would be the ratio between the total volume of RGB space (i.e. 16,777,216) and the volume of the protected ‘bubble’ in each case (which would be, for a radius of 20 units, $\frac{4}{3} \times \pi \times 20^3 = 33,510$ cubic units, i.e. 500 colours in total; or, for a radius of 10 units, 4,005 colours)

²⁵ <https://www.baeldung.com/cs/compute-similarity-of-colours>

²⁶ T. Horiuchi and S. Tominaga (2014). Color Similarity. In: ‘Computer Vision’, K. Ikeuchi, (ed.), Springer, Boston, MA. https://doi.org/10.1007/978-0-387-31439-6_450. (Available at: https://link.springer.com/referenceworkentry/10.1007/978-0-387-31439-6_450)

evolutions in the protection landscape, with a number of sound marks (including Intel's jingle, MGM's lion roar, and Netflix's 'tu-dum') already registered as brand identifiers, and the abolishment of the requirement in the EU for marks to be represented graphically, with multimedia files now permitted in applications²⁷. From the point of view of potential technological and legislative developments which may allow comparison of marks, sound marks can be 'fingerprinted' through digital analysis techniques, and it may be possible to define algorithms (perhaps incorporating elements of AI-based analysis) to measure degrees of similarity between words (see Part 2) and logos.

Conclusions – and suggestions for a possible protection framework

Colours can be key distinctive characteristics of particular brands, but currently the framework for protecting specific shades is poorly defined, and the extent of protection is unclear, in part due to a lack of definitive case law. These points mean that many registered colour marks offer protection which is not particularly robust, with the registrations often liable to third-party attempts at invalidation.

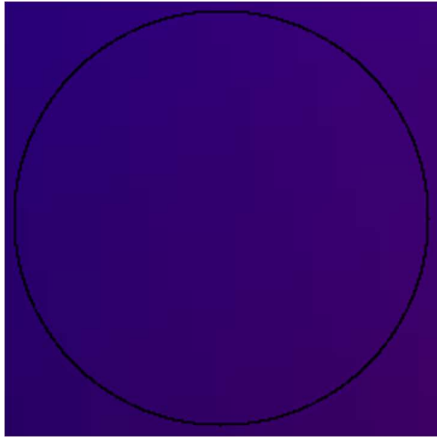
Part of the solution is a programme of consumer education on the distinctiveness of colours as brand identifiers, but it certainly seems to be the case that the legal framework could benefit from the use of rigorous mathematical descriptions, which would remove much of the subjectivity involved in mark comparisons, and allow thresholds for protection to be defined. Similar approaches could also potentially be applied to any other types of marks for which description algorithms could be produced.

As discussed above, for colour marks specifically, it would seem reasonable that the protection afforded by a particular colour-mark registration should also cover very close (but not necessarily identical) colours (for the product class(es) in question). This idea (which has analogies with the concepts of identity and similarity in regular trademarks) would circumvent the possibility of a third-party attempting to circumvent the protection by using a variant shade differing by (say) one or two RGB points. The subtlety comes in defining the exact degree of difference (i.e. the maximum colour 'distance', d) which should be covered by a colour-mark registration. A suggested value would seem to be of the order of $d = 10$ RGB units. Admittedly this would not cover all variants of (say) the Dairy Milk packaging colours shown in Figure 5, but this would seem reasonable as (for example) the colour of image (c) (approximately 30 RGB units different from the other three) does visibly appear significantly different.

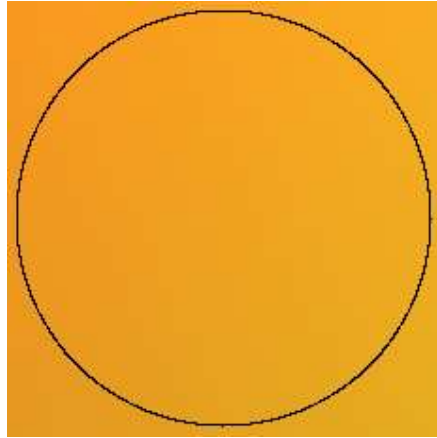
What would this suggestion look like in practice? Figure 9 shows cross sections through the suggested protected 'bubbles' of radius (d) 10 units in RGB space for six well-known colour trademarks (i.e. Figure 9(a) is equivalent to the innermost circle in Figure 8(b)). The overall effect seems reasonable; the pixels within each of the circles all visually appear nominally to be 'the same colour' as each other, whilst actually encompassing a small range of RGB variations. A value of $d = 10$ would also (as per *footnote #24*) provide a framework where approximately 4,000 distinct colours within RGB space could be protected. Most significantly, the application of a mathematical framework makes it possible to precisely define degrees of similarity or difference, and could potentially be built into a case-law structure.

²⁷ <https://www.iamstobbs.com/opinion/why-brand-owners-should-be-conscious-of-sound-trade-marks>

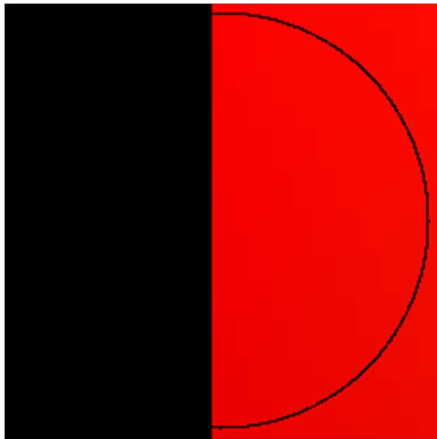
(a)



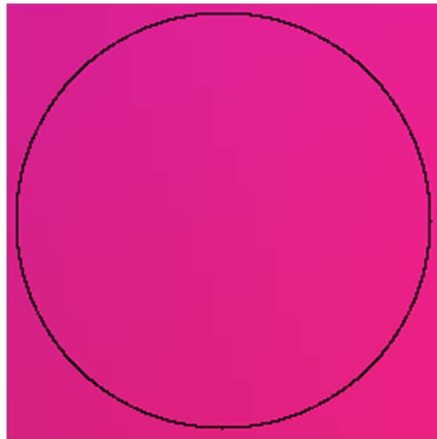
(b)



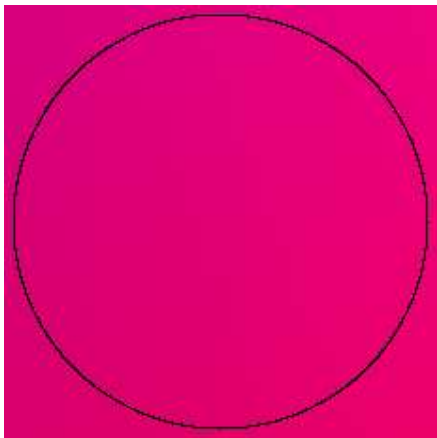
(c)



(d)



(e)



(f)

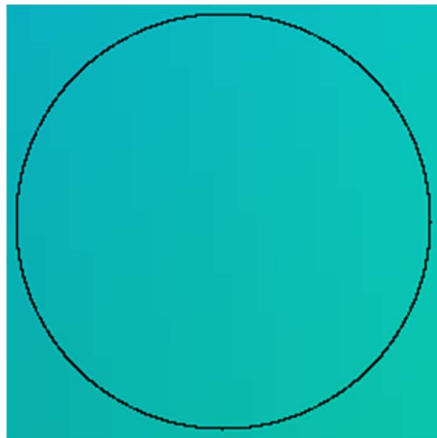


Figure 9 (*previous page*): Cross sections through suggested protected ‘bubbles’ of radius (d) 10 RGB units surrounding the following colours (which, in some cases, may be just approximations to the registered marks):

- (a) [50,0,110] – Cadbury’s
- (b) [239,162,32]²⁸ – Veuve Clicquot
- (c) [244,0,0]²⁹ (slice perpendicular to blue axis) – Coca Cola
- (d) [224,33,138]³⁰ – Barbie
- (e) [226,0,116]³¹ – T-Mobile
- (f) [10,186,181]³² – Tiffany

In each case, the slices are taken perpendicular to the axis of the *least* dominant colour in the protected mark, to illustrate the variability caused by changes in the two *most* dominant colours

Of course, this suggested threshold of 10 RGB units is based just on *my* subjective opinion of what roughly constitutes colours being ‘the same’ versus ‘different’, and it is likely that other consumers may have alternative views. Recent comments by Lord Clement-Jones, following on from the Influence at Work / Stobbs study ‘The Psychology of Lookalikes’ (*footnote #20*), have highlighted the importance of considering psychological and behavioural analyses in IP disputes, particularly in relation to brand lookalikes³³. It is likely that future research concerning the impact of colour variations on subjective perceptions of brand association (or not) will be key to a ‘ground-up’ approach for defining the thresholds which should be applied in IP protection decisions.

There is also the possibility of more complicated situations arising; for example, in cases where a trademark dispute concerns a *combination* of colours (such as the orange-and-grey schemes of Stihl / Emes), might we expect the protection to need to cover slightly greater variations of each of the colours *individually*, when considered together as a *single* mark? Furthermore, decisions regarding the degree of overlap of the associated *goods and services* classes of potentially competing products are likely to continue to add an additional degree of subjectivity.

²⁸ <https://logos-world.net/veuve-clicquot-logo/>

²⁹ <https://usbrandcolors.com/coca-cola-colors/>

³⁰ <https://encycolorpedia.com/e0218a>

³¹ <https://www.brandcolorcode.com/t-mobile>

³² <https://encycolorpedia.com/0abab5>

³³ https://www.linkedin.com/posts/geoff-steward-20404015_good-to-know-that-the-psychology-of-lookalikes-activity-7183447412542181377-6omH

Part 2: Explorations in similarity measurement for word marks

Introduction

In Part 1, I explored some initial ideas behind the objective measurement of the difference between two colours, and how it might be possible to apply this to trademark disputes relating to colour similarity. In the context of intellectual property, colours are a special case because they can be *exactly* defined (using, say, red-green-blue (RGB) specifications), and differences therefore precisely quantified (even if there still remain a number of questions regarding determination of the overlap between associated goods and services, and the definitions of thresholds relating to degrees of similarity and the umbrella of IP protection).

More generally, assessment of degrees of similarity between marks (logos, word marks, etc.) is key to a wide range of IP disputes, but is generally recognised to include a number of subjective elements. Ideally, it would be extremely useful if – at least to some degree – these measurements could more exactly be defined within an objective framework (as for colour marks), but this is extremely complex and (even considering *just* word marks, with *no* attempt to consider associated logos or imagery, classes of goods and services, or even the *meaning* of the words) there are a number of factors to consider. At the very least, any defined metrics would need to account of the following factors:

- The calculated degree of similarity between two marks differing in a particular way (say, by just one replaced character) should be considered to be lower if the marks are shorter (e.g. we might want the metric to find that ‘LG’ and ‘LV’ are less similar to each other than are (say) ‘Starbucks’ and ‘Starmucks’)
- The degree of similarity may also be dependent on the exact nature of the difference between the marks – for example, two marks differing by one just having an ‘s’ on the end might be considered to be more similar than two marks where a different letter is appended.
- A ‘one-size-fits-all’ metric might need to take account of a number of different *types* of similarity, including visual (i.e. spelling, in the case of a word mark), phonetic (i.e. pronunciation – as might be relevant to the relationship between e.g. ‘Starbucks’ and ‘Sardarbuksh’) and conceptual. (Any algorithmic assessment of conceptual similarity would need to take account of misspellings and/or homophones where the meaning is (or may be) preserved, but is not considered further in this article.)
- Any metric reflecting the overall level of threat of infringement may need to take account of the degree of commonness or distinctiveness of (elements of) one or both of the marks (e.g. what is the risk of a ‘clash’ between marks where the only overlapping element is a very common string (such as ‘McDonalds’ v ‘McSweet’, where the only overlapping element is the common string ‘Mc’), or between the quantitatively very similar words ‘Iceland’ and ‘Ireland’?).

Whilst acknowledging all of the above caveats, there are a number of well-established algorithms for quantifying the degree of similarity between two text strings (again, *not* considering the *meaning* of the word or phrase (‘lexical’ or ‘semantic’ similarity)^{34,35,36}), and it is informative to consider how effectively these are able to quantify the difference between various pairs of marks involved in previous dispute cases. As examples, we can consider:

³⁴ <https://stackoverflow.com/questions/71822208/how-do-i-calculate-similarity-between-two-words-to-detect-if-they-are-duplicates>

³⁵ <https://www.baeldung.com/cs/semantic-similarity-of-two-phrases>

³⁶ <https://www.geeksforgeeks.org/python-word-similarity-using-spacy/>

- Starbucks v Charbucks
- Starbucks v Sardarbuksh
- McDonalds v McSweet
- Louboutin v Lubov
- Louis Vuitton v Chewy Vuiton³⁷
- Puma v Coma
- Nike v Nuke
- Lakme v LikeMe
- MDH v MHS
- Mahindra v Mahendra³⁸
- Magnavox v Multivox
- HpnotiQ v Hopnotic
- Cana v Canya
- Seiko v Seycos³⁹
- Casoria v Castoria
- Trucool v Turcool
- Lucozade v Glucos-Aid
- Bacchus v Cacchus⁴⁰
- Simoniz v Permanize
- Zirco v Cozirc
- Cresco v Kresco
- Intelect v Entelec⁴¹
- Bisleri v Bilseri⁴²

Subjectively, I would consider the most similar pairings to include Mahindra/Mahendra, Casoria/Castoria and Bisleri/Bilseri (all relatively distinctive names, with the pairs just differing by one letter (replaced, added or transposed) in the *middle* of the name (rather than – say – at the start where most noticeable)), and the least similar to be McDonalds/McSweet, Simoniz/Permanize and MDH/MHS. Accordingly, we might hope that any truly effective similarity metric would reflect this.

³⁷ <https://www.tramatm.com/blog/category/other/sound-alikes-5-high-profile-trademark-disputes-involving-phonetic-similarity>

³⁸ <https://www.intepat.com/blog/deceptively-similar-trademarks-examples-case-study/>

³⁹ <https://gouchevlaw.com/likelihood-confusion-5-examples-similar-trademarks/>

⁴⁰ <https://banwo-ighodalo.com/grey-matter/trademark-infringement-analyzing-the-concept-of-confusingly-similar-trademarks>

⁴¹ <https://www.upcounsel.com/similar-trademarks-examples>

⁴² <https://www.forbesindia.com/article/news/belsri-bisleri-bilseri-bisleri-brislei/81445/1>

Standard algorithms for calculating string similarity^{43,44,45,46}

There exist a number of standard calculation approaches for quantifying the degree of similarity of two text strings, some of the most commonly used of which are outlined below.

A. Spelling-based metrics

1. Hamming distance

This is one of the simplest in a class of methods known as ‘**edit-based algorithms**’ (that is, quantifying the number of edits required to transform one string into the other). Hamming distance compares two strings (of identical length) on a character-by-character basis, to determine which pairs are the same, and which differ, giving a metric equal to the number of non-identical characters – e.g. the Hamming distance between ‘TIME’ and ‘MINE’ is 2 (with the two differences being T/M and M/N). In a refinement of the algorithm, the metric can be **normalised** by dividing it by the length of the string(s), thereby giving a representation of the *proportion* (or percentage) of the strings which are *different*. In the above case, the normalised Hamming distance is $2/4 = 0.50$; a lower value means that the two strings are more similar.

2. Levenshtein distance

This is one of the most commonly used algorithms, and quantifies the number of edits (character insertion, deletion, or substitution) necessary to transform one string into the other (and can thereby be calculated for strings of differing lengths). A greater value (i.e. a larger number of edits) means the strings are less similar (or, equivalently, a lower value means the strings are more similar). For example, the Levenshtein distance between ‘CLOCK’ and ‘CLONE’ is 2.

A modified version of this metric is the **Damerau-Levenshtein distance**, which also permits transpositions (swaps) of adjacent characters.

3. Jaro similarity⁴⁷

This metric provides a measure of similarity (between 0 and 1), with a higher value indicating strings which are more similar. It can be calculated for two strings s_1 and s_2 (with lengths $|s_1|$ and $|s_2|$, respectively), in terms of the two parameters m and t :

m is the number of matching characters, where two characters from s_1 and s_2 are deemed to ‘match’ if they are the same character and are not more than a certain distance apart (defined to be one character less than half the length of the longer string)

t is the number of transpositions; this value is equal to half the total value of instances of equivalently positioned characters within the two strings differing from each other

Then, the Jaro similarity is 0 if $m = 0$, or is otherwise equal to $\frac{1}{3} [m/|s_1| + m/|s_2| + (m-t)/m]$.

⁴³ <https://www.analyticsvidhya.com/blog/2021/02/a-simple-guide-to-metrics-for-calculating-string-similarity/>

⁴⁴ <https://medium.com/@ahmetmnirkocaman/how-to-measure-text-similarity-a-comprehensive-guide-6c6f24fc01fe>

⁴⁵ <https://yassineelkhal.medium.com/the-complete-guide-to-string-similarity-algorithms-1290ad07c6b7>

⁴⁶ https://corpustools.readthedocs.io/en/master/string_similarity.html

⁴⁷ <https://statisticaloddsandends.wordpress.com/2019/09/11/what-is-jaro-jaro-winkler-similarity/>

In a modified version of the metric, **Jaro-Winkler similarity** can be calculated from Jaro similarity, by including a weighting to take greater account of matching characters occurring at the *start* of the strings.

4. Jaccard similarity

This is an example of a '**token-based algorithm**', which calculate similarity by breaking down the strings into smaller sub-strings ('tokens') and quantifying the degree to which the sets of tokens overlap (i.e. are common to both strings). This approach can be implemented by considering the individual words in multi-word strings, or characters or groups of characters ('n-grams') in individual words. Jaccard similarity is defined as the intersection of the two sets of tokens (i.e. the number appearing in *both* strings) divided by the union of the two sets (i.e. the number of tokens in *total*).

A related metric is Sørensen-Dice similarity, in which the denominator of the metric is instead defined as the average size of the two token sets.

5. Cosine similarity

Cosine similarity can be defined for parameters which can be expressed as vectors, and can be applied to strings by expressing them in terms of word frequencies, for example. The metric is expressed as the cosine of the angle between the vectors, thereby falling in a range between -1 (entirely dissimilar – i.e. vectors pointing in entirely opposite directions) and +1 (entirely similar).

6. Ratcliff-Obershelp similarity

This is an example of a '**sequence-based algorithm**', which quantify similarity according to the closeness of sequences of characters (or tokens). Ratcliff-Obershelp similarity is one example which considers the longest common subsequence (LCS) present in both strings – i.e. a set of characters appearing in the same *order*, though not necessarily in *consecutive positions* in the strings. It is defined as twice the length of the LCS divided by the sum of the lengths of the two strings, and returns a value between 0 and 1.

B. Pronunciation-based metrics

The basic principle behind comparison of *pronunciations* is the conversion of each string to a phonetic representation, and then comparing these representations against each other, using one (or more) of the (usually edit-based) algorithms represented above⁴⁸.

The key element is therefore the production of a phonetic representation of each string, for which a number of options are available. Some of the most common are outlined in Part B of the following section.

⁴⁸ <https://ai.stackexchange.com/questions/28556/how-to-measure-the-similarity-the-pronunciation-of-two-words>

Practical implementations of similarity matching (using Python)

A. Spelling-based metrics

1. 'Fuzzy' matching

'Fuzzy' matching is a general term given to the process of identifying strings which *approximately* match a particular pattern (and, by extension, quantifying the degree of similarity between two strings which are non-identical). One convenient implementation of fuzzy matching is achieved by the Python library package '*fuzzywuzzy*'⁴⁹, which compares two strings using an algorithm based on Levenshtein distance, and returns a score between 0 and 100, indicating the degree of similarity. The simplest implementation of this algorithm uses the so-called '*ratio*' function, but there are other more complex variants, including '*partial-ratio*' (which, when comparing strings of differing lengths, matches the shorter string against each substring of the same length of the other), and '*token-sort-ratio*' and '*token-set-ratio*' (most meaningful for multi-word strings) (which split the string into tokens – i.e. distinct words – and consider similarity disregarding the *order* of the tokens, or considering the set of *unique* tokens, respectively)^{50,51}.

2. Jaro-Winkler similarity calculation

The Python library package '*Levenshtein*', which underlies '*fuzzywuzzy*', also provides implementations of a number of other algorithms, including Jaro-Winkler similarity⁵². This method may be one of the most applicable to calculation of similarity for word marks, due to its incorporation of a weighting providing an emphasis on characters towards the *start* of the strings (where, arguably, consumers may be more likely to notice differences between similar marks).

Both of the above algorithms also have the advantage that they incorporate elements of score *normalisation* – i.e. they address the first bullet point raised at the start of Part 2, accounting for the fact that equivalent differences are more 'significant' when occurring in shorter strings. For example, the two algorithms ('*fuzzywuzzy – ratio*' (*fuzz.ratio*) and '*Levenshtein – jaro_winkler*' (*Levenshtein.jaro_winkler*)) give similarity scores of 50 and 67 (respectively) when comparing 'lg' and 'lv', and 89 and 96 (respectively) when comparing 'starbucks' and 'starmucks'.

Even just considering these two algorithms, the scores produced for the pairs of marks listed previously do seem to provide meaningful measures of the (subjective!) degrees of similarity, as shown in Table 2.

⁴⁹ <https://pypi.org/project/fuzzywuzzy/>

⁵⁰ <https://marcobonanzani.com/2015/02/25/fuzzy-string-matching-in-python/>

⁵¹ <https://towardsdatascience.com/fuzzy-string-matching-in-python-68f240d910fe>

⁵² <https://rapidfuzz.github.io/Levenshtein/levenshtein.html#jaro-winkler>

Marks	<i>fuzz.ratio</i> (F_{Lev})	<i>Jaro Winkler</i> (sim_j)
starbucks/charbucks	78	85
starbucks/sardarbuksh	70	82
mcdonalds/mcsweet	38	50
louboutin/lubov	57	66
louis vuitton/chewy vuiton	56	61
puma/coma	50	67
nike/nuke	75	85
lakme/likeme	73	77
mdh/mhs	67	56
mahindra/mahendra	88	94
magnavox/multivox	50	67
hpnotiq/hopnotic	80	83
cana/canya	89	95
seiko/seycos	55	76
casoria/castoria	93	97
trucool/turcool	86	96
lucozade/glucos-aid	67	78
bacchus/cacchus	86	85
simoniz/permanize	50	67
zirco/cozirc	73	82
cresco/kresco	83	89
intelect/entelec	80	76
bisleri/bilseri	86	96

Key:

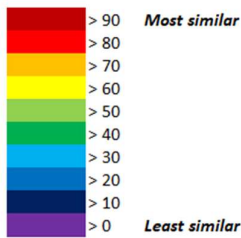


Table 2: Similarity scores for the (case-insensitive) pairs of marks, as given by the *fuzz.ratio* and *Levenshtein.jaro_winkler* algorithms (the first returns a value between 0 and 100; the second between 0 and 1, but has been multiplied by 100 so as to be directly comparable)

B. Pronunciation-based metrics

As discussed in the previous section, when considering the (comparison of) pronunciation of strings (or word marks), the first step is to create a phonetic representation of the string. There are a number of standard algorithms available for this, of which some of the most common are as follows (together with references to Python library packages offering their implementation).

- i. **IPA (International Phonetic Alphabet)**⁵³ **transcription** (available via Python package `'eng_to_ipa'`⁵⁴) – This option has the limitation that it is only available for dictionary words, and is probably therefore not generally suitable for analysis of arbitrary strings. In the IPA notation, 'starbucks' (for example) would be transcribed as **'star bəks**.
- ii. The **Soundex** algorithm provides a means of encoding a string according to its (English) pronunciation, generating a four-character output comprising a letter and three digits. The algorithm is formulated such that, essentially, the initial letter of the encoding is the first letter of the string in question, and the subsequent consonants (up to a maximum of three) are encoded with numbers, such that similarly-pronounced consonants (i.e. those articulated by a speaker in a similar way) are assigned the same digit (e.g. b, f, p and v correspond to '1' in American Soundex)⁵⁵. The *'fuzzy'* Python package⁵⁶ includes an implementation of Soundex conversion (as `fuzzy.Soundex(4)`)⁵⁷; this package provides an output for 'starbucks' of **S361**. Modifications of Soundex include Metaphone⁵⁸, which takes account of a number of phonetic inconsistencies in English, together with later versions such as Double Metaphone (also implemented in *'fuzzy'*, as `fuzzy.DMetaphone()`), which is applicable to other languages and also returns a primary and a secondary representation of each string, to account for certain ambiguous cases.
- iii. The **NYSIIS** (New York State Identification and Intelligence System) phonetic code⁵⁹ is similar to Soundex, but also incorporates a number of improvements to accuracy. Unlike Soundex, it includes a representation of the *whole* string, and also includes a number of other helpful inclusions, such as disregarding any trailing 's'. NYSIIS encoding is also implemented in *'fuzzy'*, as `fuzzy.nysiis`. This algorithm provides an output for 'starbucks' of **STARBAC**.
- iv. The **match rating approach** (MRA) is another similar algorithm which includes both an encoding process and an explicit similarity comparison⁶⁰. The output is a value less than 6, and greater than a threshold which is dependent on the length of the strings. It is implemented by the Python package *'jellyfish'*⁶¹ (as `jellyfish.match_rating_comparison`)⁶² – which also includes implementations of many of the above algorithms – though this version outputs only a 'true' or 'false' according to whether or not a match has been identified.

Table 3 shows the similarity scores for the same pairs of marks / strings as considered previously, using the phonetic Soundex and NYSIIS encodings of each of the marks, and comparing these representations using the *fuzz.ratio* algorithm.

⁵³ https://en.wikipedia.org/wiki/International_Phonetic_Alphabet

⁵⁴ <https://pypi.org/project/eng-to-ipa/>

⁵⁵ <https://en.wikipedia.org/wiki/Soundex>

⁵⁶ <https://pypi.org/project/Fuzzy/>

⁵⁷ <https://stackoverflow.com/questions/35403335/is-there-a-soundex-function-for-python>

⁵⁸ <https://en.wikipedia.org/wiki/Metaphone>

⁵⁹ https://en.wikipedia.org/wiki/New_York_State_Identification_and_Intelligence_System

⁶⁰ https://en.wikipedia.org/wiki/Match_rating_approach

⁶¹ <https://pypi.org/project/jellyfish/>

⁶² <https://manpages.debian.org/testing/python-jellyfish-doc/jellyfish.3.en.html>

Mark 1	Mark 2	Soundex			NYSIIS		
		Mark 1	Mark 2	fuzz.ratio (F _{Sou})	Mark 1	Mark 2	fuzz.ratio (F _{NYSIIS})
starbucks	charbucks	S361	C612	50	STARBAC	CARBAC	77
starbucks	sardarbuksh	S361	S636	75	STARBAC	SARDARBACS	71
mcdonalds	mcsweet	M235	M230	75	MCDANALD	MCSWAT	43
louboutin	lubov	L135	L100	50	LABATAN	LABAV	67
louis vuitton	chewy vuiton	L213	C135	50	LASVATAN	CAWAVATAN	71
puma	coma	P500	C500	75	PAN	CAN	67
nike	nuke	N200	N200	100	NAC	NAC	100
lakme	likeme	L250	L250	100	LACN	LACAN	89
mdh	mhs	M300	M200	75	MDH	MH	80
mahindra	mahendra	M536	M536	100	MANDR	MANDR	100
magnavox	multivox	M251	M431	50	MAGNAVAX	MALTAVAX	75
hpnotiq	hopnotic	H153	H153	100	HPNATAG	HAPNATAC	80
cana	canya	C500	C500	100	CAN	CAN	100
seiko	seycos	S200	S200	100	SAC	SYC	67
casoria	castoria	C260	C236	75	CASAR	CASTAR	91
trucool	turcool	T624	T624	100	TRACAL	TARCAL	83
lucozade	glucos-aid	L230	G423	50	LACASAD	GLACASAD	93
bacchus	cacchus	B200	C200	75	BAC	CAC	67
simoniz	permanize	S520	P652	50	SANAN	PARNANAS	62
zirco	cozirc	Z620	C262	50	ZARC	CASARC	60
cresco	kresco	C620	K620	75	CRASC	CRASC	100
intelect	entelec	I534	E534	75	INTALACT	ENTALAC	80
bisleri	bilseri	B246	B426	75	BASLAR	BALSAR	83

Table 3: Similarity scores for the (case-insensitive) pairs of marks, as given by the *fuzz.ratio* algorithm applied to the Soundex and NYSIIS phonetic representations of the marks

C. An overall similarity metric

Based on the above four algorithms presented in Tables 2 and 3, it is possible to calculate an overall similarity metric (S), taking into account both spelling- and pronunciation-based similarity. The simplest such implementation is given by just taking the mean of the four individual scores, though it would also be possible to apply differing weightings (w_j) to each if required, as outlined below (noting that calculating the straightforward mean is just the special case where all values of w are equal to 1).

$$S = [w_{Lev} F_{Lev} + w_j sim_j + w_{Sou} F_{Sou} + w_{NYSIIS} F_{NYSIIS}] / [w_{Lev} + w_j + w_{Sou} + w_{NYSIIS}]$$

Table 4 shows the overall similarity metric for the pairs of marks, based on the calculation of the simple mean of the four individual component scores, with the pairs of marks ranked by this score (i.e. those calculated as being most to least similar).

Marks	S
cana/canya	96
mahindra/mahendra	96
trucool/turcool	91
nike/nuke	90
casoria/castoria	89
cresco/kresco	87
hpnoti/hopnotic	86
bisleri/bilseri	85
lakme/likeme	85
bacchus/cacchus	78
intelect/entelec	78
starbucks/sardarbuksh	75
seiko/seycos	75
starbucks/charbucks	73
lucozade/glucos-aid	72
mdh/mhs	69
zirco/cozirc	66
puma/coma	65
magnavox/multivox	60
louboutin/lubov	60
louis vuitton/chewy vuiton	60
simoniz/permanize	57
mcdonalds/mcsweet	52

Table 4: Overall similarity scores (**S**) for the pairs of marks, calculated as the mean of the four individual component similarity scores

Discussion

The overall similarity metric, **S** (Table 4), seems to perform relatively well at providing an objective measurement of mark similarity, consistent with what I (subjectively!) would consider reasonable (in terms of ranking the pairs in the 'right' sort of order).

The exact details of any such formulation (in terms of which individual algorithms should be used and how their outputs should be weighted relative to each other) are certainly up for discussion, but this analysis does suggest that it should be possible to construct some sort of objective similarity metric which could be applied to IP disputes and potentially incorporated into case law. The above formulation does also already take into account many of the desired features of an overall algorithm, such as including elements of downweighting the impact of any final 's' (through the incorporation of NYSIIS) (though future formulations may be better placed by *explicitly* (further) reducing the contribution assigned to any trailing 's'), putting greater weight on similarity at the *start* of the strings (through the use of Jaro-Winkler), and utilising metrics which include normalisations relative to the length of the strings.

Of course, there are also many other additional subtleties. Should Zirco and Cozirc be considered more similar than the metric suggests, since the marks consist just of the same two syllables in a different order? (This would require some sort of explicit token-based approach.) Should other pairs of marks sharing key elements in common be deemed more similar than the metric would imply?

The answer to this question may be dependent on an assessment of how distinctive the common element is, for the relevant areas of goods and services, and it may be that more sophisticated algorithms may need to take account of such factors. This type of analysis has also taken no account of the *meanings* of marks, or of associated characteristics such as logos or fonts.

Additional factors which have not yet been addressed are the levels of prominence, distinctiveness or commonness of the mark(s). Realistically, these features are more relevant to the determination of the *likelihood of confusion* than of the *similarity* of the marks, and probably should sit within a different analytic model. A *full* assessment of potential infringement is a much more complex prospect, involving consideration of a range of factors, including real-world use.

Nevertheless, there are some relatively simple quantitative characteristics which might be relevant and are worthy of discussion here. One illustration of this point is in the Iceland/Ireland example. These words are extremely similar, and any simple word-based metric would reflect this point (actually they would be assigned an overall similarity metric, S , of 84). However, if this clash arose in an IP dispute, one might argue that these words are simply both just common words, being used with their own meanings – or are both highly established brands, rather than one attempting to passing off as the other. It might therefore be appropriate to construct a modified score, referred to in this article (for convenience) as the ‘potential infringement threat’ score (T), which – in this case – should give a lower value. This type of argument might be less relevant if *one or both* of the marks were more unusual terms (and therefore, in these cases, it might be desirable if the threat score (T) were reduced relative to the similarity score (S) by a lesser degree)).

One possible way to account for this fact would be to use the number of results returned in response to an Internet search for each mark, as a proxy for its commonness or prominence. Dividing the overall similarity score, S , by some factor, P , which is dependent on the minimum value of the number of results returned for each of the two marks would be a way of creating an infringement threat score (T) which is more greatly reduced if both marks are common terms. One way of doing this would be to calculate the reduction factor by taking the logarithm of the number of search engine results (N_s) (so that 1,000 results would give a value of 3; 10,000 results a value of 4; 100,000 results a value of 5, and so on). This might reduce the similarity score by ‘too great’ a degree⁶³, so it might be preferable to apply an additional scaling factor, k ⁶⁴, such that:

$$T = S \cdot k / \log_{10}(\min(N_{s_brand1}, N_{s_brand2}))$$

How would this apply to the Iceland/Ireland case? This is illustrated by the calculation below.

$S = 84$	(as referenced above)
$N_{s_iceland} = 828,000,000$	(i.e. the number of results returned in response to a Google search ⁶⁵ (for example) for ‘iceland’) ⁶⁶
$N_{s_ireland} = 2,760,000,000$	(i.e. the number of results returned in response to a Google search (for example) for ‘ireland’)

⁶³ The absolute values are, of course, essentially arbitrary, but it might be desirable to formulate the algorithm to continue to output values in an approximate range between 0 and 100. In any case, the value of k would need to be fixed, if the similarity score for any given pair of marks is to be comparable against the scores for other pairs.

⁶⁴ In this formulation, therefore: $P = \log_{10}(\min(N_{s_brand1}, N_{s_brand2})) / k$, such that $T = S / P$

⁶⁵ In these cases, it is advisable to search for the brand name in quote marks, to avoid the search engine presenting results for *similar* marks

⁶⁶ As of 29-Jul-2024

$$\therefore \min(N_{s_brand1}, N_{s_brand2}) = 828,000,000$$

$$\therefore \log_{10}(\min(N_{s_brand1}, N_{s_brand2})) = 8.9$$

$$k = 2 \text{ (say)}$$

$$\therefore \underline{T = 19}$$

Applying this concept to the pairs of brands shown previously would result in potential infringement threat scores as shown in Table 5.

Marks	T
zirco/cozirc	80
trucool/turcool	66
seiko/seycos	50
bacchus/cacchus	48
bisleri/bilseri	45
simoniz/permanize	40
lucozade/glucos-aid	40
hpnatiq/hopnotic	39
starbucks/sardarbuksh	39
intelect/entelec	36
cresco/kresco	36
cana/canya	32
starbucks/charbucks	30
casoria/castoria	30
lakme/likeme	26
mahindra/mahendra	25
louis vuitton/chewy vuiton	25
magnavox/multivox	24
nike/nuke	24
mcdonalds/mcsweet	23
mdh/mhs	19
louboutin/lubov	18
puma/coma	15

Table 5: Potential infringement threat scores (**T**) for the pairs of marks, to take account of levels of brand prominence or commonness (using **k** = 2)

There are a number of key differences from the previous rankings (Table 4). The first category is a set of increases in infringement threat scores (relative to the similarity scores) for pairs of brands where one of the brands has a low online prominence (e.g. 'cozirc' (45 results), 'turcool' (599), 'permanize' (751), 'seycos' (962)), suggesting that its use is more likely to have been driven purely in reference to the more well-known brand, rather than having a well-established presence in its own right. The other main change are drops in score for pairs where *both* marks are common or non-distinctive terms (e.g. puma/coma, nike/nuke, mdh/mhs) – which again 'seems' reasonable. There are, however, some anomalies, such as the relatively low score for casoria/castoria (a result of the fact that both terms are relatively common online), which 'feels' too low, so this particular metric

may require some further 'tuning' (not least also because there are almost certainly other factors to be considered in any determination of likelihood of confusion).

Given these and the other previously discussed factors, it seems unavoidable that IP disputes will continue to involve significant degrees of subjectivity, and the case law is likely to need to evolve in order to also take account of these points. Nevertheless, some sort of objective measurements of similarity – as presented in this article – may be a step in the right direction at offering the potential for an increase in levels of consistency across legal decisions.

Acknowledgements

This article was inspired by a discussion at Stobbs CaseFest #15 (London, 11-Jul-2024) – with thanks to Emma Pettipher, Jessica Wolff, Will Haig, Richard Ferguson, John Weston, Geoff Weller, Jacob Larking, Chris Sleep and others for their input.