Journal of Dentistry 91 (2019) 103244

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent

A yellowness index for use in dentistry

Chelsea Sullivan[a], Qianqian Pan[a], Stephen Westland[a,*], Roger Ellwood[b]

[a] School of Design, University of Leeds, Leeds, LS2 9JT, United Kingdom
[b] Colgate Fellow, Colgate Palmolive Dental Health Unit, Williams House, Manchester Science Park, Manchester, M15 6SE, United Kingdom

ARTICLE INFO

Keywords:
Tooth color
Whiteness
Yellowness
Tooth whitening

ABSTRACT

Objectives: To develop an equation that predicts the perceptual yellowness of teeth.
Methods: A large set of new psychophysical yellowness data were generated from an experiment where 500 participants each ranked a set of 58 shade guide samples. Two existing equations (WIO and b*) and one new equation (YIO) were evaluated by comparing their values for the 58 shade guide tabs with the psychophysical data. Coefficient of determination (r^2), '% wrong decisions', and STRESS were used as measures of performance. The YIO equation was optimized using these data to maximize the r^2 value. A validation set of psychophysical data was prepared in an experiment where 40 participants each ranked 5 sets of 9 samples that were viewed on an emissive display. The candidate equations were evaluated using these data and the r^2, %WD, and STRESS metrics.
Results: All three metrics YIO, WIO and b* were strongly correlated with perceptual yellowness. YIO and WIO both showed stronger correlation than b*.
Conclusions: A new yellowness equation YIO has been developed to correlate with tooth yellowness. It is suggested that tooth yellowness and whiteness are highly related concepts.

1. Introduction

There has been increasing interest in how to assess the color of teeth as modern dentistry has moved from treatment of dental caries and oral health to include tooth aesthetics and greater concern for how patients evaluate aesthetic outcomes [1,2]. A number of studies have revealed that many people are dissatisfied with the color of their teeth [3]. The social reasons for this satisfaction are becoming more evident as it is now understood that whitened teeth can lead to judgements that are more positive on personality traits such as social competence and appeal, intellectual ability and relationship satisfaction [2]. The market for tooth whitening has consequently grown significantly and various tooth-whitening approaches and products are now available [4,5] as the mechanisms of whitening have become more widely understood [6]. The last couple of decades have seen increased efforts to measure the color of teeth objectively using an internationally recognized system of colorimetry published by the Commission Internationale l'Éclairage (CIE) [7] where the measurements are obtained by various technologies including colorimeters, spectrophotometers, spectroradiometers and cameras [8]. A number of different color spaces exist but the most common color spaces in dental research are CIE XYZ and CIELAB. Although CIELAB is more widely used in dentistry, the CIE XYZ space is also important, in particular because of the associated chromaticity space defined by the chromaticity coordinates x and y.

It is not easy to relate three-dimensional changes in CIE XYZ or CIELAB values to changes in a univariate perceptual attribute such as whiteness. Although some studies [9] have found that the CIELAB b* parameter is correlated with perceptual whiteness (allowing the other two dimensions of color space to be ignored) this approach is unlikely to be valid generally. This problem received substantial attention from other industries long before it became a concern in dentistry since whiteness is an important attribute for many products [10,11]. A number of equations have consequently been developed for the assessment of whiteness in the textile and paper industries, most notably the CIE Whiteness index WIC (Eqn. 1) [12,13]:

$$WIC = Y + 800(x_n - x) + 1700(y_n - y) \quad (1)$$

where Y, x and y are the colorimetric properties (luminance and chromaticity values) of the sample to be assessed and x_n, y_n are the chromaticity values of the reference white (usually the light source used to view the samples). Two equations have been optimized specifically to assess the whiteness of teeth. The WIO equation (Eqn. 2) has the same form as the WIC equation but the coefficients were optimized to best fit experimental data on the perception of tooth whiteness [14,15]:

* Corresponding author at: School of Design, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, United Kingdom.
E-mail address: s.westland@leeds.ac.uk (S. Westland).

https://doi.org/10.1016/j.jdent.2019.103244
Received 25 March 2019; Received in revised form 7 November 2019; Accepted 11 November 2019
0300-5712/

$$WIO = Y + 1075.012(x_n\text{-}x) + 145.516(y_n\text{-}y). \quad (2)$$

Recently, a new equation WI_D [16] has been developed for use in dentistry that is based upon the CIELAB color space: Thus,

$$WI_D = 0.511L^*\text{-}2.24a^*\text{-}1.100b^* \quad (3)$$

A recent cross-cultural study of perceptual whiteness concluded that the concept of whiteness is consistent between different cultures, genders and age groups [17]. This suggests that a single whiteness equation could be used to measure whiteness perception for these different groups. However, whiteness is not the only colour perceptual attribute of teeth that is of concern to dentists and patients. Some studies have reported changes in lightness and yellowness and have used changes in the CIELAB parameters L* and b* to correlate with perceptual changes in lightness and yellowness respectively [18]. It is far from clear that b* can generally be used as a correlate of perceptual yellowness since neither of the cartesian coordinates a* and b* can indicate hue on their own. Consequently, some studies have used yellowness indices to assess perceptual yellowness [19]. However, there is no yellowness index that has been shown to be effective at correlating with perceptual yellowness of teeth.

More generally, it is far from clear what the relationship is between the percepts of whiteness and yellowness in the context of teeth. Generally perceptual whiteness might be expected to increase as perceptual yellowness decreases. But to what extent are perceptual whiteness and yellowness antonyms, particularly within the constraints of tooth color space? In order to address this question and to develop a yellowness index for use in dentistry, a psychophysical scaling experiment was conducted to measure visual yellowness for a set of samples. A yellowness index was developed based on these psychophysical data and was validated using data from a second experiment where participants viewed digital simulations of teeth on a color-calibrated display.

2. Methodology

2.1. Psychophysical experiments

In a previous study of perceptual whiteness a psychophysical experiment was carried out where participants were asked to rank a set of shade guide tabs in order of whiteness [17]. In the same experiment the participants were also asked to rank the shade guide tabs in order of yellowness and it is these data that are now reported. At each of 5 distinct geographical locations (these being UK, India, Brazil, USA and China), 100 participants were recruited in a balanced design to allow the effect of age, gender and culture to be assessed (25 young males, 25 young females, 25 old males and 25 old females). The young group was aged 18–30 and the old group was aged 30-60. Each participant was asked to rank each of 58 shade guide tabs in order of yellowness in a viewing cabinet under D65 illumination. The set of shade guide tabs consisted of 29 Vita Toothguide 3D-Master tabs and 29 custom-made Vita Extended Bleachedguide 3D-Master tabs. A neutral grey card (with CIE L* of approximately 50) was provided for each geographical center to cover the interior base of the viewing cabinet so that the same background would be used in each study (in the UK the cabinet was provided by Verivide Ltd, UK). The spectral reflectance factors of each tab were measured using a Konica-Minolta CM-2600d reflectance spectrophotometer and subsequently converted to CIE XYZ values for D65 illuminant (1931 CIE observer). Each of the 500 participants was asked to place the 58 samples in rank order of decreasing perceptual yellowness. These rank orders were analyzed to generate interval scale yellowness values [20]; that is, for each tab a number was derived that represents the relative perceptual yellowness of that tab. These interval-scale data will be used to optimize an equation that can predict perceptual yellowness.

A set of validation data were also collected based on five related psychophysical experiments. For each of the five experiments 9 digitally simulated teeth were displayed on a color-calibrated display. In each case the 9 samples were derived from a color center. The CIELAB values for the five color centers are displayed in Table 1. These CIELAB values were chosen to be approximately evenly spaced along the central axis of the gamut of tooth color in CIELAB space. The CIELAB values of the 8 additional samples (in each set of 9) were approximately at the vertices of a cube arranged around the color centre (the faces of the cube were 2 CIELAB units away from the centre in each of the L*, a* and b* directions).

Table 1
Measured CIELAB values of the five color centres used in the digital simulation.

L*	a*	b*
81.03	−2.11	0.27
74.31	−2.06	3.16
70.35	−2.01	6.92
68.81	−2.18	11.58
65.29	−1.24	15.38

Each of 40 participants viewed each set of 9 samples (one set at a time) on a digital display (from a distance of approximately 50 cm) in a darkened room and was asked to rank the 9 samples in order of decreasing yellowness. The participants did this by moving the samples on the screen using the mouse in a graphical-user interface written using MATLAB software. The rank orders were recorded for each participant and for each set and were used, as before, to calculate interval scale yellowness values for the tabs in each set. The teeth samples were displayed on a neutral grey background with an L* value of approximately 50.

2.2. Assessment of yellowness indices

Three methods are used to quantify the agreement between the candidate yellowness indices and the psychophysically derived interval scale yellowness values. The coefficient of determination and the % wrong decision were used because they have been used in related studies and will allow easy comparison. [14–16]. The r^2 value between two sets of data is the coefficient of determination and this is the square of the Pearson correlation coefficient r. Values closer to one indicate a high correlation between the two sets of data. The % wrong-decisions criterion is obtained by comparing each sample in a data set with each other and calculating the number of times that the yellowness metric would disagree about which one of a pair is the yellowest compared with the average visual decision of the whole group of observers (denoted by the visual scale values). A yellowness metric agrees with the visual data if the % wrong decisions value is low. The standardised residual sum of squares (STRESS) is also used and is defined thus: [21,22].

$$STRESS = 100[(\Sigma(Y_i\text{-}FV_i)^2)/(\Sigma(F^2V_i)^2)]^{0.5} \quad (4)$$

where Y_i and V_I are the computed and visual scale values respectively for each of the i samples and F is a factor that adjusts the scales of Y_I and V_I and is given by $(\Sigma Y_i^2)/(\Sigma Y_i V_i)$. STRESS is always in the range 0–100 and greater values represent worse agreement between visual and computed yellowness values. The squared ratio of STRESS values from two ways of computing yellowness indices follows an F-distribution and this can be used to determine whether the two models are statistically significantly different or not (in this study a 95 % confidence interval is used). STRESS is introduced in this work as a way to differentiate between the performances of the different models that are used.

Three candidate yellowness indices are considered in this study: (a) WIO (Eqn. 2),(b) b* and (c) a new yellowness equation with the same form as Eqn. 2. The second of these, b*, is included because some clinical studies have used b* as a measure of tooth-color change during a tooth-whitening procedure. Note that normally, for WIO, higher

Table 2
Performance of the candidate equations on the 58-tab data that was derived from 500 participants.

Candidate metric	r^2	% WD	STRESS
WIO	0.97	4.84	27.01
b*	0.91	9.68	94.26
YIO	0.97	4.96	30.35

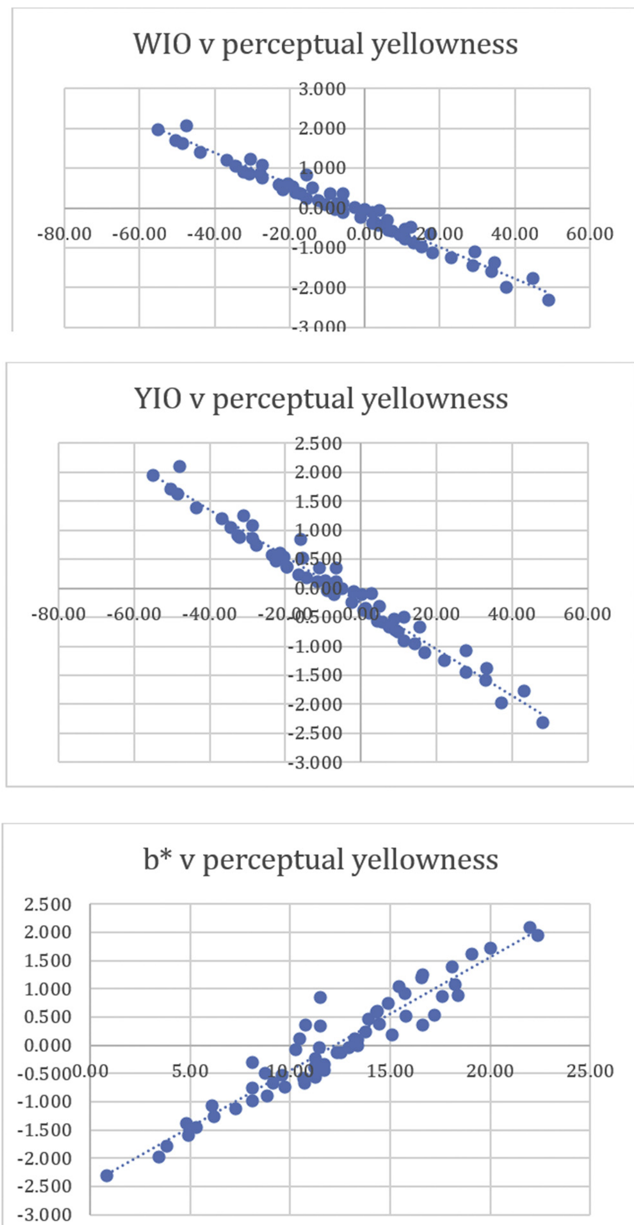

Fig. 1. Correlation between the three candidate metrics and the perceptual yellowness data for the 58 samples. Data are shown for WIO (upper), YIO (middle) and b* (lower) with r2 values of 0.97, 0.97 and 0.91 respectively.

values represent increased whiteness; in this study, lower values of WIO will be used to represent increased "yellowness". The new yellowness equation is based on the generic form shown in Eqn. 5. However, the coefficients p and q were optimized so as to maximize the r^2 value between the yellowness score YIO and the perceptual yellowness values for the 58 samples that were ranked by 500 participants. The optimization was performed using the Solver algorithm in Microsoft Excel.

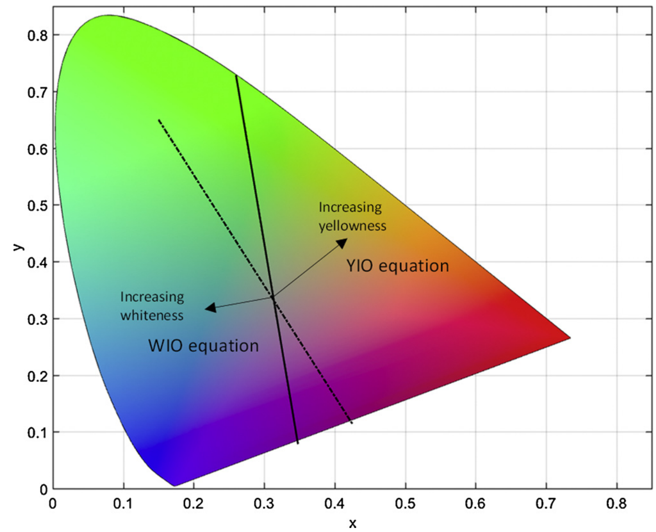

Fig. 2. Iso-whiteness (solid line) and iso-yellowness (dashed line) for the WIO and YIO equations respectively.

Table 3
Performance (r^2) of the candidate equations on the five sets of digitally simulated data from 40 participants.

Candidate metric	Set 1	Set 2	Set 3	Set 4	Set 5
WIO	0.97	0.89	0.92	0.95	0.94
b*	0.55	0.38	0.70	0.57	0.58
YIO	0.95	0.88	0.95	0.98	0.97

Table 4
Performance (% WD) of the candidate equations on the five sets of digitally simulated data from 40 participants.

Candidate metric	Set 1	Set 2	Set 3	Set 4	Set 5
WIO	2.8	2.8	5.6	2.8	2.8
b*	38.9	38.9	22.2	33.3	22.2
YIO	5.5	11.1	2.8	5.6	5.6

Table 5
Performance (STRESS) of the candidate equations on the five sets of digitally simulated data from 40 participants.

Candidate metric	Set 1	Set 2	Set 3	Set 4	Set 5
WIO	99.83	99.68	99.68	99.05	96.45
b*	89.77	79.75	94.68	99.05	99.45
YIO	99.82	99.66	99.60	98.73	93.99

This algorithm is a form of generalised reduced gradient descent [23].

$$YIO = Y + p(x_n - x) + q(y_n - y). \quad (5)$$

3. Results

Table 2 shows the coefficient of determination r^2, % wrong-decisions, and STRESS values for the three candidate yellowness indices on the training set of 58 samples. The squared STRESS ratios are 9.64 (for YIO vs. b*), 12.18 (for WIO vs. b*) and 1.26 (for WIO vs. YIO). The critical F value for F_c (0.975, 57, 57) is 1.67 which gives a confidence interval of [0.56–1.67]; ratios outside of this confidence intervals indicate statistically different performance. Therefore the performance of b* is significantly worse than for either YIO or WIO; however, the WIO and YIO formulae are statistically indistinguishable.

The YIO equation was optimized to maximize r^2 for these data and the optimized equation that resulted is shown as Eqn. 6.

$$YIO = -Y - 851.716(x_n - x) - 436.962(y_n - y). \quad (6)$$

Fig. 1 shows the correlations between each of the three candidate equations and the perceptual yellowness values for the 58 samples.

Fig. 2 graphically illustrates the difference between the WIO and YIO equations using the concept of iso-whiteness and iso-yellowness lines in a CIE chromaticity diagram. In this diagram all of the colours along the dashed line have the same perceptual yellowness (as predicted by Eq. 5). A set of imaginary lines (parallel to the dashed line) would also have iso-yellowness. However, yellowness increases perpendicular to these lines towards the yellow region of colour space as denoted by the arrow. Fig. 2 shows that increasing whiteness according to WIO is from an orange to cyan direction whereas increasing yellowness according to YIO is from a blue to yellow equation. This would suggest that although tooth whiteness and tooth yellowness are highly related concepts, yellowness is not simply the antonym of whiteness.

Table 3 shows the r^2 values for the three candidate metrics with the perceptual yellowness data for the five sets of digitally-simulated tooth samples. Table 4 shows the %WD for the same samples. The b* metric produces lower r^2 values and a higher number of wrong decisions than either the WIO or YIO index. Table 5 shows the STRESS data for these samples. It is less clear from these data that there is a difference between the three metrics. The critical F value for F_c (0.975, 8, 8) is 4.43 which gives a confidence interval of [0.23–4.43]. For all five data sets, none of the squared STRESS ratios fall outside of this confidence interval for any of the metric comparisons. Therefore according to the STRESS analysis the three metrics are statistically identical.

4. Discussion

All three metrics show good correlation with perceptual yellowness. The r^2 values for WIO and YIO for the 58 physical shade guide tabs are identical and are both higher than for the b* metric. In terms of percent wrong decisions, the WIO and YIO equations both performed similarly to each other and better than b*. STRESS analysis confirmed that the performance of b* was significantly worse than for either WIO or YIO.

These results suggest that perceptual yellowness of teeth is not simply a correlate of CIELAB b*. The strong performance of WIO relative to YIO is perhaps unexpected given that YIO was developed on this test set and the performance would therefore have been expected to be optimized.

The performance of the three equations on the digitally simulated tooth samples is weaker than for the 58 physical tabs. The reason for this is that the colorimetric values of the physical tabs are more highly correlated. Changes in color between these tabs move along the gamut of natural tooth color changes (for example, b* always decreases as L* increases). This reduces the degrees of freedom in the data and enables the equations to more easily fit the data. The digitally simulated samples constitute a much more rigorous test for the equations. For these samples, as b* decreases, L* may decrease, increase or remain constant.

For the digitally simulated samples, although all three metrics perform well, for both r^2 values and percent wrong decisions WIO and YIO perform better than b*. For percent wrong decisions there is a suggestion that WIO even though it is based on perceptual whiteness may perform slightly better than YIO. However, STRESS analysis revealed no statistical difference between the three metrics; this may be because the sample size ($n = 9$) was so small for each of the 5 sets of data.

It is suggested from Fig. 2 that tooth yellowness may not be a perfect antonym of tooth whiteness. However, these data suggest that both WIO and YIO could be used to reflect perceptual yellowness. It is likely that perceptual yellowness is a less universally agreed construct than whiteness which is reflected in the agreement between participants. Further work may be needed to understand the concept of yellowness in tooth color space.

5. Conclusion

This study reports two new psychophysical experiments to assess perceptual yellowness of two sets of samples. The perceptual yellowness samples derived from these samples were used to derive a new yellowness equation YIO that is optimized for tooth color. The whiteness equation WIO and b* were also evaluated as predictors of perceptual yellowness. All three metrics YIO, WIO and b* were strongly correlated with perceptual yellowness and might be used as measures of perceptual yellowness.

Declaration of Competing Interest

Part of this work was carried through financial support from the Colgate Palmolive Company. Dr Ellwood is an employee of the Colgate Palmolive Company. The work also used previously published data that was also carried out with the financial support of the Colgate Palmolive Company.

Acknowledgements

The work was mainly funded by Colgate Palmolive Company (USA) and the internal grant number at the University of Leeds was 113533.

References

[1] A. Hasegawa, A. Motonomi, I. Ikeda, S. Kawaguchi, Color of natural tooth crown in Japanese people, Color Res. Appl. 25 (1) (2000) 43–48.
[2] A. Joiner, W. Luo, Tooth colour and whiteness: a review, J. Dent. 67S (2017) S3–S10.
[3] A. Joiner, I. Hopkinson, Y. Deng, S. Westland, A review of tooth colour and whiteness, J. Dent. 36S (2008) S2–S7.
[4] R.W. Gerlach, Whitening paradigms 1 year later: introduction of a novel professional tooth-bleaching system, Compend. Contin. Educ. Dent. 23 (2002) 4–8.
[5] A. Joiner, Whitening toothpastes: a review of the literature, J. Dent. 38S (2010) e17–e24.
[6] S.R. Kwon, P.W. Wertz, Review of the mechanism of tooth whitening, J. Esthet. Restor. Dent. 27 (2015) 240–257.
[7] J. Schanda (Ed.), Colorimetry: Understanding the CIE System, John Wiley & Sons, 2007.
[8] Q. Pan, S. Westland, Tooth color and whitening - digital technologies, J. Dent. 74S (2018) S42–S46.
[9] R.W. Gerlach, M.L. Barker, P.A. Sagel, Objective and subjective whitening response of two self-directed bleaching systems, Am. J. Dent. 15 (2002) 7A–12A.
[10] R.S. Berns, Billmeyer and Saltzman's Principles of Color Technology, Wiley-Blackwell, 2019.
[11] S. Westland, M.R. Luo (Ed.), CIE Whiteness, in Encyclopedia of Color Science and Technology, Springer, 2015, , https://doi.org/10.1007/978-3-642-27851-8_5-1 1-5.
[12] E. Ganz, Whiteness formulas: a selection, Appl. Opt. 18 (7) (1979) 1073–1078.
[13] ASTM, Designation E313-73: Standard Test Method for Indices of Whiteness and Yellowness of Near-white, Opaque Materials, (1993).
[14] W. Luo, S. Westland, P. Brunton, R. Ellwood, I.A. Pretty, N. Mohan, Comparison of the ability of different colour indices to assess changes in tooth whiteness, J. Dent. 35 (2) (2007) 109–116.
[15] W. Luo, S. Westland, R. Ellwood, I. Pretty, V. Cheung, Development of a whiteness index for dentistry, J. Dent. 37 (Suppl 1) (2009) e21–e26, https://doi.org/10.1016/j.jdent.2009.05.011.
[16] M. del Mar Pérez, R. Ghinea, M.J. Rivas, A. Yebra, A.M. Ionescu, R.D. Paravina, L.J. Herrera, Development of a customized whiteness index for dentistry based on CIELAB color space, Dent. Mater. 32 (3) (2016) 461–467.
[17] Q. Pan, S. Westland, R. Ellwood, Evaluation of colorimetric indices for the assessment of tooth whiteness, J. Dent. 76 (2018) 132–136.
[18] S.T. Kim, P.V. Abbott, P. McGinley, The effects of Ledermix paste on discolouration of mature teeth, Int. Endod. J. 33 (3) (2001) 227–232.
[19] L. Rubio, J.M. Sioli, J. Suarez, M.J. Gaitan, S. Martin-de-las-Heras, Spectrophotometric analysis of color changes in teeth incinerated at increasing temperatures, Forensic Sci. Int. 252 (193) (2015) e1–193 e6.
[20] G.R. Dowling, D.F. Midgley, Using rank values as an interval scale, Psychol. Mark. 8 (1) (1991) 37–41.
[21] M. Melgosa, R. Huertas, R.S. Berns, Performance of recent advanced color-difference formulas using the standardized residual sum of squares index, J. Opt. Soc. Am. A 25 (7) (2008) 1828–1834.
[22] P.A. García, R. Huertas, M. Melgosa, G. Cui, Measurement of the relationship between perceived and computed color differences, J. Opt. Soc. Am. A 24 (7) (2007) 1823–1829.
[23] RAIRO, Oper. Res. 8 (V3) (1974) 73–103.