The Hourglass Model Revisited

Yosephine Susanto
Nanyang Technological University

Andrew G. Livingstone
University of Exeter

Bee Chin Ng

Erik Cambria
Nanyang Technological University

***Abstract*—Recent developments in the field of AI have fostered multidisciplinary research in various disciplines, including computer science, linguistics, and psychology. Intelligence, in fact, is much more than just IQ: it comprises many other kinds of intelligence, including physical intelligence, cultural intelligence, linguistic intelligence, and EQ. While traditional classification tasks and standard phenomena in computer science are easy to define, however, emotions are still a rather mysterious subject of study. That is why so many different emotion classifications have been proposed in the literature and there is still no common agreement on a universal emotion categorization model. In this work, we revisit the Hourglass of Emotions, an emotion categorization model optimized for polarity detection, based on some recent empirical evidence in the context of sentiment analysis. This new model does not claim to offer the ultimate emotion categorization but it proves the most effective for the task of sentiment analysis.**

IN 1872, Charles Darwin was one of the first scientists to argue that all humans, and even animals, show emotions through remarkably similar behaviors [1]. Since then, there has been broad consensus on how and why emotions have evolved in most creatures. The definition and the categorization of emotions, however, have always been a big challenge for the research community [2], [3]. To date, in fact, there are still active debates on whether some basic emotions, e.g., *surprise* [4], should be defined as emotions at all. In this work, we do not aim to initiate any new philosophical discussion on emotions nor to propose the ultimate emotion categorization model. Our goal is simply to review some of the most popular emotion models in the context of computer science and, hence, propose a new version of the Hourglass of Emotions [5], a categorization model for concept-level sentiment analysis.

  Published by the IEEE Computer Society

The remainder of this paper is organized as follows: next section discusses the main emotion models proposed in the literature; later, the revised version of the Hourglass model is presented in detail; then, an evaluation of the model on three sentiment analysis datasets is provided; finally, the last section offers concluding remarks.

RELATED WORK

Emotion research has increased significantly over the past few years thanks to the recent developments in the field of AI. The question, in fact, is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any [6]. One of the earliest efforts in developing an emotion model was made by Shaver et al. [7]. They first selected a group of words and had them classified as emotion words and non-emotion words. This step resulted in 135 emotion words, which were then annotated based on their similarity and grouped into categories so that inter-category similarity was minimized but intra-category similarity maximized. Using the typical prototyping approach, they managed to develop an abstract-to-concrete emotion hierarchy and discovered six emotions on the hierarchy's lowest level: *joy*, *love*, *surprise*, *sadness*, *anger*, and *fear*. This emotion study implied that most emotions are fuzzy or indistinct and they are combinations of these six basic emotions, which cannot be further divided.

Later, Ortony and Turner argued against the view that basic emotions are psychologically primitive [8]. They proposed that all emotions are discrete, independent, and related to each other through a hierarchical structure, hence there is no basic set of emotions that serve as the constituents of others. Having refuted the existence of basic emotions, Ortony, Clore, and Collins introduced their own emotion model (termed OCC from the initials of the three authors) [9]. The OCC model classifies emotions into 22 emotion types. The hierarchy contains three branches, namely consequences of events (e.g., pleased or displeased), actions of agents (e.g., approving or disapproving), and aspects of objects (e.g., liking or disliking). A number of ambiguities of the emotions defined in the OCC model were later identified and discussed by Steunebrink et al. [10], who extended the model to 24 emotion categories.

A few years after the original OCC model was proposed, Mehrabian proposed the Valence/Arousal model [11], a popular model in psychology that places specific emotion concepts in a circumflex model of core affect defined by two basic dimensions: Arousal, which ranges from high to low, and Valence, which varies from positive to negative. Another very popular model, based on facial expressions, was later proposed by Ekman [12]. The model only consists of six emotions (*anger*, *fear*, *disgust*, *joy*, *sadness*, and *surprise*) but turned out to be one of the most used models in the literature for its simplicity and applicability. Many subsequent models are based on Ekman's model, e.g., Plutchik's wheel of emotions [13]. Likewise, the Hourglass of Emotions [5] is a reinterpretation of Plutchik's model for sentiment analysis. Many more models have been proposed in the literature [14], mostly to adapt previous models to different disciplines, modalities, or applications.

THE REVISITED MODEL

After almost a decade of using the Hourglass model [5] in the context of sentiment analysis, we realized that this presents several issues, namely:

- uncanny color associations;
- presence of neutral emotions;
- absence of some polar emotions;
- wrong association of antithetic emotions;
- low polarity scores for compound emotions;
- absence of self-conscious or moral emotions.

Uncanny color associations

While this was not a matter that affected the accuracy of sentiment analysis, it has been a pressing issue for a while since many researchers in the community questioned the choice of some colors of the Hourglass, e.g., blue for *surprise*, green for *fear*, and purple for both *sadness* and *disgust*. In line with recent studies on the association between colors and emotions [15], we assigned tendentially warm colors to positive emotions and cold colors to negative ones (Figure 1). This also ensures a better distinction between different emotions (e.g., *sadness* and *disgust* are now blue and green, respectively) and an enhanced organization of the model (positive emotions now reside in the upper part of the Hourglass while negative ones are at the bottom).

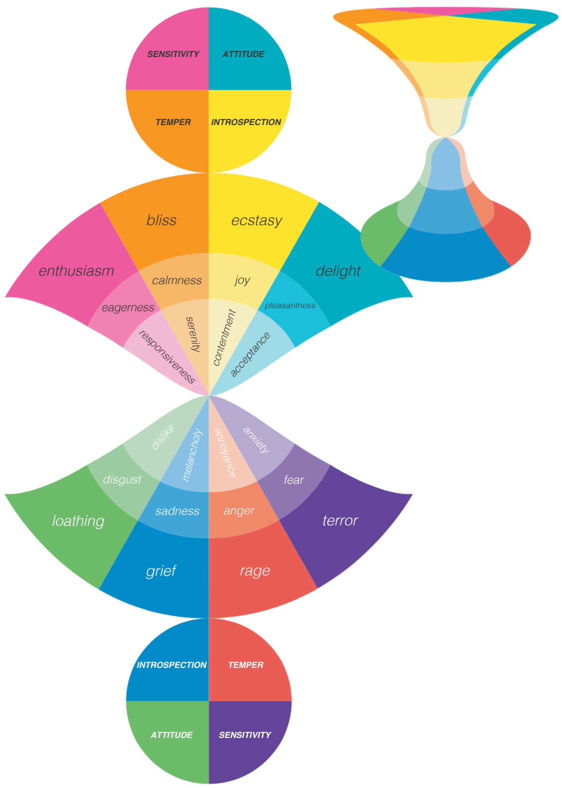

Figure 1. The Hourglass model revisited.

Presence of neutral emotions

One of the main problems with the previous model was the presence of ambiguous emotions (e.g., *distraction* [16]) and, especially, neutral emotions, e.g., *surprise*. Here, we do not want to debate whether *surprise* is an emotion or not [4] but we definitely do not want it in a model that is catered for sentiment analysis as this will lead to the wrong categorization of all concepts (words and multi-word expressions) that are semantically associated with it. *Surprise*, in fact, only becomes polar when coupled with positive or negative emotions (Table 1).

Absence of some polar emotions

Another issue with the original model was the absence of some important polar emotions, e.g., *calmness* and *eagerness*. All the concepts associated with such emotions, e.g., `deep_breath` or `volunteer`, were going undetected by the model and, hence, miscategorized as neutral. This issue extended to germane emotions, e.g., *enthusiasm* and *bliss*, and concepts associated with them, e.g., `ambition` or `meditation`.

Wrong association of antithetic emotions

One of the main advantages of having an emotion categorization model is to be able to classify unknown concepts based on known features. For example, if the model did not contain the emotion *discomfort*, it could look up its opposite (*comfort*) and flip its polarity to obtain the polarity of the unknown concept. This mechanism works well in the new model, as emotions are now organized with respect to their polarity (Table 2), but it generated a lot of errors in the previous version of the Hourglass, as this contained wrong associations of antithetic emotions, e.g., *anger* and *fear* (which are both negative) or *surprise* and *anticipation* (which are opposite in terms of meaning but not in terms of polarity).

Low polarity scores for compound emotions

The main goal of sentiment analysis is to calculate the polarity value (positive or negative) of a piece of text, an image or a video. In many applications, polarity intensity also plays an important role for classification and decision-making. The old Hourglass model had a big shortcoming in this sense: to make sure the polarity value stayed between -1 (extreme negativity) and +1 (extreme positivity), a static normalization factor was introduced. Such a normalization factor, however, made the polarity intensity of most concepts very low. Concepts with high intensity were not the ones with high emotional charge but rather those that were associated with compound emotions (e.g., *hatred*) because of more dimensions active at the same time (e.g., *anger* and *fear*).

JOY	PLEASANTNESS	love	enjoyment	amusement
	EAGERNESS	euphoria	excitement	thrill
	CALMNESS	enlightenment	relaxation	sweet idleness
SADNESS	DISGUST	hate	guilt	remorse
	FEAR	distress	troubledness	misery
	ANGER	envy	bitterness	resentment
CALMNESS	PLEASANTNESS	assertiveness	compassion	empathy
	EAGERNESS	focus	determination	perseverance
	FEAR	carelessness	laxity	looseness
ANGER	DISGUST	hatred	ruthlessness	viciousness
	FEAR	nastiness	coercion	possessiveness
	EAGERNESS	stubborness	obstinacy	mulishness
PLEASANTNESS	DISGUST	shamelessness	cheekiness	brazenness
	EAGERNESS	kindness	audacity	hospitality
	FEAR	awe	submission	reverence
DISGUST	JOY	morbidness	schadenfreude	gloat
	FEAR	impiety	cowardness	inhospitality
	EAGERNESS	recklessness	temerity	rashness
EXPECTATION	JOY	hope	anticipation	optimism
	SADNESS	hopelessness	despair	pessimism
	EAGERNESS	vigilance	alertness	caution
SURPRISE	ANGER	shock	outrage	thunderstruckness
	FEAR	alarm	dismay	dumbstruckness
	PLEASANTNESS	amazement	astonishment	wonderstruckness

Table 1. Examples of compound emotions.

INTROSPECTION					
ECSTASY	JOY	CONTENTMENT	MELANCHOLY	SADNESS	GRIEF
elation	happiness	satisfaction	pensiveness	unhappiness	desperation
jubilation	cheerfulness	gratification	abandonment	sorrow	gloom
exultation	joviality	fulfilment	emptiness	dejection	depression
glee	gaiety	light-heartedness	down-heartedness	heavy-heartedness	broken-heartedness
felicity	high-spiritedness	frivolity	nostalgia	low-spiritedness	woe
TEMPER					
BLISS	CALMNESS	SERENITY	ANNOYANCE	ANGER	RAGE
placidity	tranquillity	quietude	disquietude	vexation	fury
peacefulness	equanimity	comfort	discomfort	exasperation	wrath
beatitude	composure	ease	unease	aggressiveness	ferocity
gladness	restfulness	imperturbability	perturbability	madness	enragement
relief	soothingness	carefreeness	frustration	acrimoniousness	vengeance
ATTITUDE					
DELIGHT	PLEASANTNESS	ACCEPTANCE	DISLIKE	DISGUST	LOATHING
admiration	appreciation	approval	disapproval	disappointment	contempt
adoration	fondness	favorability	distaste	detestation	revulsion
glorification	predilection	propensity	rejection	disdain	scorn
devotion	respect	belief	disbelief	disrespect	repugnance
enthrallment	trust	worthiness	worthlessness	distrust	abhorrence
SENSITIVITY					
ENTHUSIASM	EAGERNESS	RESPONSIVENESS	ANXIETY	FEAR	TERROR
zeal	keenness	decisiveness	indecisiveness	fright	horror
zest	willingness	receptiveness	apprehension	dread	panic
passion	motivation	agreeableness	helplessness	trepidation	appalment
avidity	inspiration	approachableness	agitation	angst	petrification
fervor	dedication	amenability	discouragement	scare	aghastness

Table 2. New emotion classification with five sample emotion words for each category.

To this end, we replaced the old normalization factor with a new dynamic quantity that is directly proportional to the number of active dimensions:

$$p_c = \frac{I_c + T_c + A_c + S_c}{|sgn(I_c)| + |sgn(T_c)| + |sgn(A_c)| + |sgn(S_c)|} \quad (1)$$

where c is an input concept, p is the polarity value of such concept, I is the value of Introspection (the *joy*-versus-*sadness* dimension), T is the value of Temper (the *calmness*-versus-*anger* dimension), A is the value of Attitude (the *pleasantness*-versus-*disgust* dimension), and S is the value of Sensitivity (the *eagerness*-versus-*fear* dimension). Before, a negative concept (e.g., `death`) associated with a strong emotion (e.g., *grief*) would not result in a high (negative) polarity because its affective intensity would have been divided by 3. Now, that same intensity remains intact because the denominator of the polarity formula is equal to 1, since only one dimension (Introspection) is active. The denominator will actually be equal to 1 for most concepts, as most concepts are only associated with one emotion; it will be equal to 2 for concepts that are associated with bidimensional emotions like *love* (*joy+pleasantness*) and *submission* (*fear+pleasantness*); it will be equal to 3 for those few concepts that are associated with tridimensional emotions like *bittersweetness* (*sadness+anger+pleasantness*); finally, it will be 4 for those very rare concepts that are associated with compound emotions that span all dimensions like *jealousy* (*anger+fear+sadness+disgust*).

Absence of self-conscious or moral emotions

The old Hourglass model systematically excluded what are commonly known as self-conscious or moral emotions such as *pride*, *prejudice*, *guilt*, *shame*, *embarrassment* or *humiliation*. This has been a serious issue as it caused the model to be unable to recognize this pretty large subset of emotions and, hence, the polarity (and the concepts) associated with them. We solved this issue by encapsulating such emotions as subdimensions of Attitude (Table 3).

Emotions like *pride* and *confidence*, in fact, can be interpreted as positive Attitude (*pleasantness* and *acceptance*, respectively) directed at oneself. Likewise, *embarrassment* and *guilt* rep-

ATTITUDE (toward self)					
DELIGHT	PLEASANTNESS	ACCEPTANCE	DISLIKE	DISGUST	LOATHING
self-respect	pride	confidence	low-confidence	shame	self-contempt
self-adoration	self-appreciation	security	insecurity	self-blame	self-loathing
self-devotion	self-attraction	modesty	embarrassment	self-disgust	self-abasement
self-regard	self-formation	self-esteem	low-self-esteem	disgrace	self-denigration
self-fulfilment	self-motivation	assurance	self-deprecation	self-pity	self-condemnation
ATTITUDE (toward others)					
DELIGHT	PLEASANTNESS	ACCEPTANCE	DISLIKE	DISGUST	LOATHING
morality	sociability	sympathy	antipathy	asociability	immorality
generosity	appeasement	fairness	unfairness	greed	malevolence
self-sacrifice	affability	humbleness	prejudice	meanness	turpitude
magnanimity	conviviality	humility	hostility	humiliation	wickedness
bounty	friendliness	gratitude	ingratitude	unfriendliness	xenophobia

Table 3. The subdimensions of Attitude with five sample emotion words per category.

resent negative Attitude (*dislike* and *disgust*, respectively) directed at oneself. Similarly, *magnanimity* and *sociability* can be considered positive Attitude (*delight* and *pleasantness*, respectively) towards others, while *humiliation* and *malevolence* represent negative Attitude (*disgust* and *loathing*, respectively) towards others.

EVALUATION

We tested the new Hourglass model against some of the above-mentioned emotion categorization models on three sentiment benchmarks: the Blitzer Dataset [17], the Movie Review Dataset [18], and the Amazon dataset [19]. The first consists of product reviews in seven different domains and contains 3,800 positive sentences and 3,410 negative ones. The second is about movie reviews and is composed of 4,800 positive sentences and 4,813 negative ones. Finally, the Amazon dataset contains the reviews of 453 mobile phones, which were split into sentences and labeled as positive, neutral, or negative. The final dataset contains 48,680 negative sentences and 64,121 positive ones.

We used these three datasets to compare how the new Hourglass model performs on the task of polarity detection in comparison with the models proposed by Shaver [7], Ekman [12], Plutchik [13], the OCC models [9], [10], and the previous Hourglass model [5] (Table 4). For this experiment, we considered sentiment analysis as a binary classification problem (positive versus negative) and, hence, we left out models that focus more on intensity, e.g., the Valence/Arousal model.

The evaluation was performed by connecting the concepts of SenticNet [20], a commonsense knowledge base for sentiment analysis, to a positive or negative polarity via the emotions of each model and by using sentic patterns [19] to calculate the polarity of each sentence in the datasets. Sentic patterns model sentences as electronic circuits: sentiment words are ‘sources’ while other words are ‘elements’, e.g., `very` is an amplifier, `not` is a logical complement, `rather` is a resistor, `but` is an OR-like element that gives preference to one of its inputs (Figure 2). Thus, for each emotion model, a polarity was firstly assigned to each concept encountered in a sentence based on its connections with positive or negative emotions in the graph of SenticNet and, secondly, sentic patterns were used to calculate the final polarity of the sentence.

As expected, the accuracy of text sentiment analysis using the models of Ekman and Shaver is low as both are based on facial expressions and, hence, cover a very limited set of emotions. Ekman’s model, in particular, is not very good for detecting polarity from text because, unlike Shaver’s model, it is unbalanced (as it consists of 2 positive emotions and 4 negative ones).

Model	Blitzer dataset	Pang&Lee dataset	Amazon dataset
Ekman's model	66.87%	65.92%	59.53%
Shaver's model	67.12%	66.73%	60.89%
Plutchik's model	86.94%	85.79%	80.91%
Hourglass model	88.27%	88.12%	82.75%
OCC model	89.15%	88.73%	84.76%
OCC model revisited	90.41%	89.41%	85.93%
Hourglass model revisited	**94.72%**	**93.29%**	**89.85%**

Table 4. Comparison of emotion models on three datasets for sentiment analysis.

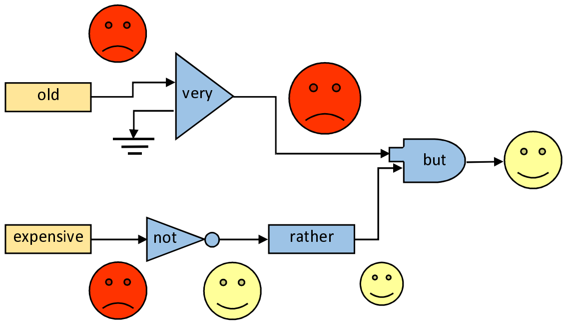

Figure 2. Sentiment data flow for the sentence "The car is very old but rather not expensive" via sentic patterns.

Plutchik's model and the old Hourglass model performed better since they both cover 24 emotions (plus compound emotions), but still suffer from the presence of neutral emotions and the absence of some important polar emotions. The categorization of *surprise* as a positive emotion, in particular, caused a lot of misclassifications because (at least in the context of sentiment analysis from product reviews) it is more often associated with negative emotions, e.g., *shock*. The old Hourglass model performed slightly better because it covers 8 additional compound emotions that are particularly useful for polarity detection from product reviews, e.g., *frustration*.

The OCC models performed considerably better thanks to the absence of *surprise* and the presence of some moral emotions that turned out to be important for sentiment analysis, e.g., *regret* (as in unhappy customers regretting having bought a product). The revisited model performed slightly better than the original thanks to the addition of *interest* and *disgust*.

Finally, the Hourglass model revisited is the best-performing model thanks to the better interpretation of neutral emotions like *surprise* and *expectation* and their combination with other polar emotions (Table 1), the presence of important emotions like *eagerness* and *calmness* that were missing from all other models (Table 2), and the inclusion of some moral emotions, e.g., *pride* and *shame*, which were missing from the previous model but are important for sentiment analysis (Table 3). Most of the misclassified sentences were using sarcasm or contained phrases with untriggered sentic patterns.

CONCLUSION

Affective neuroscience and twin disciplines have clearly demonstrated how emotions and intelligence are strictly connected. Some prominent researchers have also questioned the possibility of emulating intelligence without taking emotions into account. Emotions, however, are rather elusive entities and, hence, are difficult to categorize.

In this paper, we reviewed major emotion models and proposed a new version of the Hourglass model, a biologically-inspired and psychologically-motivated emotion categorization model for sentiment analysis.

This model represents affective states both through labels and through four independent but concomitant affective dimensions, which can potentially describe the full range of emotional experiences that are rooted in any of us. The new version of the model provides a better color representation of emotions; it excludes neutral emotions (e.g., *surprise*) and includes some important polar emotions that were previously missing (including self-conscious and moral emotions); it better categorizes emotions in order to ensure that antithetic emotions are mirrored; finally, it calculates the polarity associated with natural language concepts with higher accuracy.

In the future, we plan to test the validity of the new Hourglass model on different domains (beyond product reviews) and different modalities (beyond text). We also plan to develop mechanisms to dynamically customize the model according to different cultures, personalities, age group, sex, and user preferences.

ACKNOWLEDGEMENTS

The authors would like to thank Prof Andrew Ortony for the insightful discussions on emotion categorization (especially on the inclusion of *calmness* and the exclusion of *surprise*) and Jocelyn, Victor and Alvin Choong for their support in designing the new graphics of the model.

REFERENCES

1. C. Darwin, *The Expression of the Emotions in Man and Animals*, John Murray, 1872.
2. J. Panksepp, "The basics of basic emotion," P. Ekman and R. J. Davidson, eds., The nature of emotion: Fundamental questions, Oxford University Press, 1994, pp. 20–24.

3. L. F. Barrett, "Are emotions natural kinds?" *Perspectives on Psychological Science*, vol. 1, no. 1, 2006, pp. 28–58.
4. A. Ortony, "Some 'basic' emotions are not emotions at all: Surprise!" G. L. Schiewer, J. Altarriba, and B. C. Ng, eds., Language and emotion: An international handbook, De Gruyter Mouton, 2021.
5. E. Cambria, A. Livingstone, and A. Hussain, "The Hourglass of Emotions," A. Esposito et al., eds., Cognitive Behavioral Systems, Lecture Notes in Computer Science, vol. 7403, Springer, Heidelberg, 2012, pp. 144–157.
6. M. Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, Simon & Schuster, New York, 2006.
7. P. Shaver et al., "Emotion knowledge: further exploration of a prototype approach," *Journal of Personality and Social Psychology*, vol. 52, no. 6, 1987, pp. 1061–1086.
8. A. Ortony and T. Turner, "What's basic about basic emotions?" *Psychological Review*, vol. 97, no. 3, 1990, pp. 315–331.
9. A. Ortony, G. Clore, and A. Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge, 1988.
10. B. Steunebrink, M. Dastani, and J. Meyer, "The OCC model revisited," *Workshop on Emotion and Computing*, 2009.
11. A. Mehrabian, "Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, 1996, pp. 261–292.
12. P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, 1992, pp. 169–200.
13. R. Plutchik, "The nature of emotions," *American Scientist*, vol. 89, no. 4, 2001, pp. 344–350.
14. Z. Wang, S. Ho, and E. Cambria, "A review of emotion sensing: Categorization models and algorithms," *Multimedia Tools and Applications*, 2020.
15. F. Takahashi and Y. Kawabata, "The association between colors and emotions for emotional words and facial expressions," *Color Research and Application*, vol. 43, no. 2, 2018, pp. 247–257.
16. A. Iordan and F. Dolcos, "Brain activity and network interactions linked to valence-related differences in the impact of emotional distraction," *Cerebral Cortex*, vol. 27, no. 1, 2017, pp. 731–749.
17. J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," *ACL*, 2007, pp. 440–447.
18. B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," *ACL*, 2005, pp. 115–124.
19. S. Poria et al., "Sentiment data flow analysis by means of dynamic linguistic patterns," *IEEE Computational Intelligence Magazine*, vol. 10, no. 4, 2015, pp. 26–36.
20. E. Cambria et al., "SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis," *CIKM*, 2020, pp. 105–114.

Yosephine Susanto is a PhD student at Nanyang Technological University. Her research interests include affective computing and multilingual sentiment analysis. Susanto received a master in applied linguistics from Atma Jaya Catholic University of Indonesia. Contact her at yosephin001@e.ntu.edu.sg.

Andrew G. Livingstone is a senior lecturer at University of Exeter. His research interests lie in social identity, emotion, group processes and intergroup relations. Livingstone was awarded a PhD in social psychology from University of Exeter. Contact him at a.livingstone@exeter.ac.uk.

Bee Chin Ng is an associate professor at Nanyang Technological University. Her research interests include psycholinguistics and sociolinguistics aspects of language acquisition in multilingual contexts. Ng received a PhD in linguistics from La Trobe University. Contact her at mbcng@ntu.edu.sg.

Erik Cambria is the corresponding author and an associate professor at Nanyang Technological University. His main research interests are AI and affective computing. Cambria earned his PhD in computing science and mathematics through a joint programme between the University of Stirling and MIT Media Lab. Contact him at cambria@ntu.edu.sg.