Queueing Systems Vol. 100
Special Issue: 100 views on queues

Integrating Emotional Load into Service Operations

Galit Yom-Tov · Anat Rafaeli

February 6, 2022

Emotions are clearly important. They drive our behavior and how we behave drives our relationships, with family, friends and coworkers. Emotions are influential both in the social world and in the professional world. In particular, service systems, like contact centers and healthcare organizations, contain ample emotions and emotional expressions [9]. We suggest that emotions impact service operations and therefore that queueing models used to optimize service operations should include considerations of emotions. We describe in this note a concept that we developed with the help of many others, termed *emotional load*.

Emotions in service systems are embedded in customer behavior (e.g., anger about waiting, anxiety due to uncertainty) as well as in agent behavior (e.g., apologies for wrongs that were done or empathy and compassion to hearten customers). Emotions are integral to the everyday life of service agents, who are frequently the punching bag of the service industry because they represent the service end of (usually large and complex) operations. Organizational Behavior (OB) research shows that customer emotions are consequential. For example, people's ability to solve problems is hampered by encounters with customer anger, as is employees' burnout and turnover [8]. We take this line of work from the human behavior domain to the operational domain.

Emotional load connects emotions and operations. In [3], we propose an analogy to the operational concept of *offered load*, which is defined as $R = \lambda E[S]$, where λ is the arrival rate of customers into a system, and $E[S]$ is the expected amount of resources needed to process the customer, measured by the expected service time [7]. The analogous conceptualization of *emotional load* posits the arrival rate of events posing emotional demands (analogous to λ) and the effort that each emotional demand requires from an agent (analogous to $E[S]$) [3]. The effort element of emotional load can take multiple forms. For example, our work has shown extended verbal communication and response time [1] and emotional responsiveness [2] are forms of effort. There are prob-

Galit Yom-Tov & Anat Rafaeli
Faculty of Industrial Engineering and Management, Technion—Israel Institute of technology, Israel. E-mail: {gality, anatr}@technion.ac.il

ably other forms of effort, and all forms can likely appear independently or together. The aggregate of all effort creates the total emotional load imposed on a service agent at any given point.

We pose two main research questions to the community: (a) what are the operational implications of emotional load? (b) how should considerations of emotional load be incorporated into operational models and decisions? We have started answering these questions and describe here in short the approach or tools we used, some of our results, and the challenges we find to be still open.

We build on developments in Natural Language Processing (NLP) to monitor emotions in textual services using Sentiment Analysis (e.g., [11]). This allows the development of operational models that take sentiment information into account, which can support both retrospective analysis and real time monitoring of emotional load. For example, based on [2], we operationalized emotional load in *contact centers* using automated assessments of customer sentiment. Then, in [1] we used econometric models to measure the impact of "arrivals" of customer sentiment on agent productivity. Our analyses show that adding a negative word to a customer sentence increases agent response time by 19.7%. This results in 15.7 minutes handling time for a "negative" customer compared to 7.6 minutes to handle a "positive" customer. The effect of emotional load is 2.66 times larger than that of multitasking and an order of magnitude larger than the effects of load—core factors for the Behavioral Operations and Operations Research communities [5]. Three main issues are still open: i) what drives the operational implications of emotional load? additional agent effort (additional text that agents wrote) only partially explained the longer response time. Further investigation is needed to unravel other causes. ii) how can emotional load be measured in domains lacking textual representation of the service interactions, for example healthcare service? iii) what other events or situations create emotional load?

Our results further suggest that emotional load should be incorporated into operational models for staffing, concurrency and routing. We started incorporating emotional load into routing decisions [3,4]. Yom-Tov et al. [11] show a clear pattern of change in customer emotions during service conversations, which suggests that customer sentiment can insinuate the stage of a conversation; more negative emotions likely indicate a service conversation that is just beginning. Combining this with findings of Altman et al. [1] suggests we may predict the residual service time of an active conversation. [1] offered this as a foundation for *Sentiment-Based Routing*, namely routing customers to an agent based on the sentiment of each agent's concurrent customers. Theoretical analyses and implementation of such ideas are still open.

In [4] we extended this idea by incorporating emotion and behavioral aspects into service duration models and used them to improve routing. Specifically, this work developed dynamic models of service duration—using Hawkes processes—incorporating the mutual dependencies of customers and agents, and proved that such models fit data much better than classic service models. The models were used to predict residual service times that were then incorporated into routing algorithms. These models significantly reduced customer waiting time without need for additional resources.

We note that the contact center systems we studied include concurrency, with agents serving multiple customers in parallel. Using process sharing models, Tolga et al. [10] proved that routing customers according to the 'lowest number of customers' rule, is asymptotically optimal. Yet, simulation studies show the idea of prediction-based routing, sketched in [4], to be significantly more efficient. Hence, we call for theoretical analyses to unravel its wider potential. In particular, [4] suggests routing customers to agents predicted to have the lowest amount of total load. This routing policy replaces the currently common load indicator (number of customers agents serve), with an *estimate* of agents' future load based on customer and agent behavior. An open question is how to optimize the use of such prediction models.

Additional research showed that behavioral aspects, such as overload-based service slowdown, change the dynamics of queueing models—possibly turning a stable system into unstable—and influence staffing and blocking procedures [6]. This may also be the case for emotional load, because [1] proved a vicious cycle of reciprocal impact between emotional load and operational load. Namely, waiting increases customer negative emotions which in turn cause slower agent reaction. Hence, another open direction is how taking into account the interdependence between operational and emotional loads influences system dynamics and in turn staffing recommendations.

In short, our work shows both the theoretical significance and the operational potential of incorporating emotional load into service operation models. We use this forum to call the community to join us in this exciting journey.

References

1. D. Altman, G. B. Yom-Tov, S. Ashtar, M. O. Olivares, and A. Rafaeli. Do customer emotions affect agent speed? an empirical study of emotional load in online customer contact centers. *Manufacturing & Service Operations Management*, 23(4):854–875, 2021.
2. S. Ashtar, G. B. Yom-Tov, N. Akiva, and A. Rafaeli. When do service employees smile? response-dependent emotion regulation in emotional labor. *J. Organisational Behavior*, 42:1202–1227, 2021.
3. N. Carmeli. *Data-Based Resource-View of Service Networks: Performance Analysis, Delay Prediction and Asymptotics*. PhD thesis, Technion—Israel Institute of Technology, 2020.
4. A. Daw, A. Castellanos, G. B. Yom-Tov, J. Pender, and L. Gruendlinger. The co-production of service: Modeling service times in contact centers using Hawkes processes. Working paper, 2021.
5. M. Delasay, A. Ingolfssonb, B. Kolfal, and K. Schultz. Load effect on service times. *Eur. J. Oper. Res.*, 279:673–686, 2019.
6. J. Dong, P. Feldman, and G. B. Yom-Tov. Service system with slowdowns: Potential failures and proposed solutions. *Operations Research*, 63(2):305–324, 2015.
7. G. Eick, W. W. Massey, and W. Whitt. The physics of the M_t/G/∞ queue. *Operations Research*, 41(4):731–742, 1993.
8. A. Rafaeli, A. Erez, S. Ravid, R. Derfler-Rozin, D. Efrat-Treister, and R. Scheyer. When customers exhibit verbal aggression, employees pay cognitive costs. *J. App. Psychology*, 97(5):931–950, 2012.
9. A. Rafaeli, G. B. Yom-Tov, S. Ashtar, and D. Altman. Opportunities, tools and new insights: Evidence on emotions in service from analyses of digital traces data. In C. E. J. Härtel, W. J. Zerbe, and N. M. Ashkanasy, editors, *Emotions and Service in the Digital Age (Research on Emotions in Organizations, Vol 16)*, pages 105–136. Emerald Publishing Limited, UK, 2020.
10. T. Tezcan and J. Zhang. Routing and staffing in customer service chat systems with impatient customers. *Operations Research*, 62(4):943–956, 2014.
11. G. B. Yom-Tov, S. Ashtar, D. Altman, M. Natapov, N. Barkay, M. Westphal, and A. Rafaeli. Customer sentiment in web-based service interactions: Automated analyses and new insights. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27*, page 8 pages, NY, USA, 2018. ACM.