Unsafe Social Welfare Systems: A System Safety Analysis of the Dutch Childcare Benefit Scandal*

Maurus Enbergs[1,*], Sem J.J. Nouws[1] and Roel I.J. Dobbe[1]

[1]Delft University of Technology, Jaffalaan 5, 2628BX Delft, The Netherlands

Abstract

The introduction of algorithmic decision-making in the social welfare domain has contributed to the emergence and amplification of harm to citizens. In several cases, the use of algorithmic decision-support systems has led to the formation of so-called "digital cages"; administrative exclusion by digital information architectures. To date we still lack comprehensive theoretical concepts to describe and analyze the systemic hazards introduced by algorithmic systems. Our study illustrates the affordances of system-theoretic concepts and methods, drawing on system safety, to understand and analyze algorithmically induced hazards in public governance. We show, on example of the *Dutch Childcare Benefit Scandal*, that the system safety discipline offers powerful concepts and tools to understand, prevent, and address algorithmically induced system hazards in social welfare. By applying system safety concepts, our study contributes to the development of sociotechnical assessment approaches for algorithmic systems in public governance.

Keywords

system safety, algorithmic decision-making, systems theoretic process analysis, childcare benefit scandal

The Problem: Algorithmic interventions in the public domain often lack appreciation for the inherent complexity of the larger sociotechnical system landscape they are embedded into. In several cases, this has led to citizens being harmed, for example, through the emergence of "digital cages"; large-scale administrative exclusion by digital information architectures. The *sociotechnical dynamics* that drive the emergence of phenomena, such as digital cages, have yet to be fully understood [1]. Despite the value of empirical and theoretical analysis, we lack comprehensive theoretical concepts and methods to understand and address the harmful outcomes of algorithmic systems in their sociotechnical context.

Our Approach: By applying *systems theoretic process analysis* (STPA) to the Dutch *Childcare Benefit Scandal*, we present a first-of-its-kind demonstration and exploration of system safety analysis to understand the role of algorithmic decision-making in harms emerging in social welfare systems. System safety emphasizes an integral lens that looks at how different systemic factors contribute to unsafe outcomes, rather than focusing on individual aspects of a system [2]. It has dealt with complex safety-critical systems subject to forms of software-based automation for many decades, but its lessons have yet to be absorbed in the domain of algorithmic systems for public administration [3].

Our Findings: In our analysis, we focused on core processes in the Dutch Tax Administration during the time of the Childcare Benefit Scandal. Our STPA analysis identified three central mechanisms promoting system hazards that are identical to three known canonical safety problems:

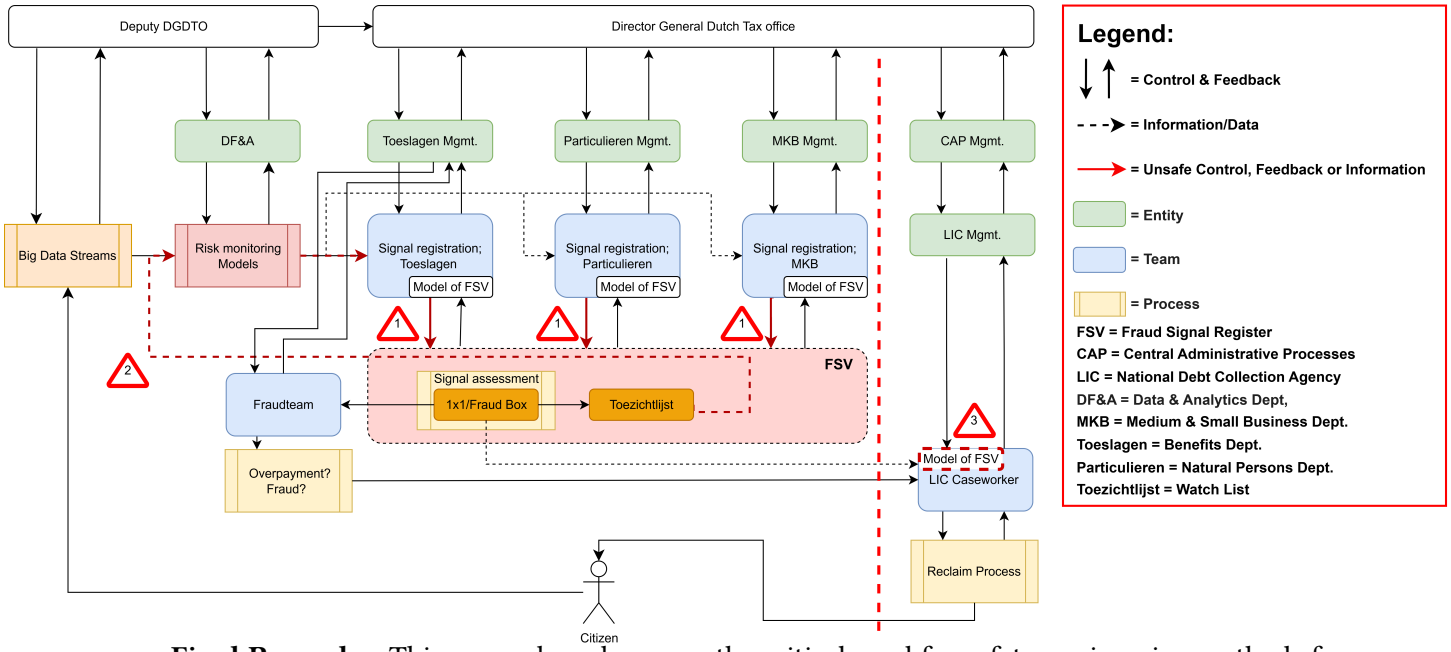
1. Lack of Control Coordination: System hazards are found in areas of overlapping controls. Such overlapping zones are situations where the same function is realized by several subsystems. During the Child Care Benefit scandal, the Tax Authority allowed several divisions to register and update risk signals on the central risk signal database using different process standards. This created a problem of control coordination between the divisions, in turn obstructing caseworkers from understanding who registered, changed, or acted upon a signal. This lack of cohesion enabled the asynchronous evolution of the processes of each division, inadvertently reducing the data quality with detrimental effects for any subsequent decisions made upon these signals.

2. Runaway feedback loops; are a known cause of failure within complex systems. Runaway loops are a system dynamic in which a process amplifies its own effects, leading to an acceleration pattern [2]. Within the Dutch Tax Administration, such a loop was caused by the missing control over the

Proceedings EGOV-CeDEM-ePart conference, August 31 - September 4, 2025, University for Continuing Education, Krems, Austria.
*Corresponding author.
m.e.enbergs-1@tudelft.nl (M. Enbergs); s.j.j.Nouws@tudelft.nl (S. J.J. Nouws); r.i.j.dobbe@tudelft.nl (R. I.J. Dobbe)
0009-0007-6840-7666 (M. Enbergs); 0000-0003-0623-6930 (S. J.J. Nouws); 0000-0003-4633-7023 (R. I.J. Dobbe)

"toezichtslijst" (monitoring list). The list consisted of citizens who were suspected of fraud or gross negligence. Once a person was on this list, any signal associated with them would be flagged by one of several big data feed risk models and forwarded to case workers for personal control. Furthermore, until 2019, data from the list was used to train and test these risk models. This closed-loop process meant that registration on the list significantly increased scrutiny and the chances of (incorrect) determination of fraud. The average time spent on the list was 4,8 years [4].

3. Mental Model Misalignment; is a function of the lack of information and feedback exchange between system components [2]. Our example refers to the misalignment of aspects embedded into the mental model of human controllers. Specifically, this involved individuals who had unknowingly submitted a faulty tax return. By default, these cases were marked with the "1x1" checkbox prior to manual assessment. However, caseworkers at the national debt collection center (LIC) had been misinformed about the implied conventions of the signal. They believed that applications with the "1x1" box ticket were verified fraud cases, eliminating the right to a personalized payment plan [4]. In reality, the "1x1" box did not determine definitive guilt. Rather, it was a mechanism by which caseworkers could indicate that, for a specific individual, further assessment was needed [4].



Final Remarks: This research underscores the critical need for safety engineering methods for algorithmic interventions in social welfare. System Safety provides practitioners with a structured framework to navigate the complexities of public algorithmic systems. By illustrating how well-understood safety problems manifested in the Childcare Benefit Scandal, we demonstrate STPA's effectiveness in identifying and addressing complex system interactions in digital social welfare systems.

Declaration on Generative AI

During the preparation of this work, the author(s) used **Grammarly** to: Grammar and spelling check, Paraphrase, and reword. Additionally, a RAG pipeline utilizing **OpenAIAPI-GPT4** was developed to query relevant documents related to the case study. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] A. Rizk, I. Lindgren, Automated Decision-Making in the Public Sector: A Multidisciplinary Literature Review, in: M. Janssen, J. Crompvoets, J. R. Gil-Garcia, H. Lee, I. Lindgren, A. Nikiforova, G. Viale Pereira (Eds.), Electronic Government, Springer Nature Switzerland, Cham, 2024, pp. 237–253. doi:10.1007/978-3-031-70274-7_15.

[2] N. Leveson, Engineering a safer world: systems thinking applied to safety, Engineering systems, MIT Press, Cambridge, Mass, 2011. OCLC: ocn719429220.

[3] R. I. J. Dobbe, System Safety and Artifcial Intelligence (2022) 15. URL: https://academic-oup-com.tudelft.idm.oclc.org/edited-volume/41989/chapter-abstract/377785597?redirectedFrom=fulltext. doi:https://doi-org.tudelft.idm.oclc.org/10.1093/oxfordhb/9780197579329.001.0001.

[4] PwC, Onderzoek effecten FSV Toeslagen - Rapport - Rijksoverheid.nl, 2021. URL: https://www.rijksoverheid.nl/documenten/rapporten/2021/12/03/onderzoek-pwc-effecten-fsv-toeslagen, last Modified: 2022-03-14T11:12 Publisher: Ministerie van Algemene Zaken.