Contrary to the classical theory of operations management, recent case studies in retail, call centers, and healthcare indicate that pooling queues may not necessarily result in less expected work in process. In this paper, we propose that this phenomenon may arise when servers are work averse and have some discretion over their choice of service capacity. We distinguish two types of work aversion, namely workload aversion and busyness aversion, and show that dedicated configurations yield less expected work in process than pooled configurations when servers exhibit high degrees of workload aversion or low degrees of busyness aversion. We also find that busyness aversion tends to hurt more to the point that it could negate the operational benefits of queue pooling at their highest potential. Overall, our work suggests that service system designers may need to consider the servers’ type and extent of work aversion as well as their degree of capacity choice discretion before pooling their workload. [Joint work with Mor Armony and Hummy Song]