Tanjo.ai
400 West Weaver Street
Carrboro, NC, 27514

How TAPs Are Created / Simulation Whitepaper

Richard Boyd

What are TAPs?

A **T**anjo **A**nimated **P**ersona (TAP) is a synthetic model of human interests, values and choices. These synthetic models, or personas, are representative of individuals, but not tied to any specific individual, and therefore protective of privacy (GDPR compliant). By modeling attitudes and interests and permitting simulations to test ideas, TAPs offer a compelling new approach to market research.

Our offering is grounded in two theses:

Thesis #1: Humans have distinct personalities, each with unique interest and value models.

Thesis #2: By leveraging data, we can simulate these interests and values and allow experimentation and testing with them that is as valid as – and in some cases superior to – results from focus groups and surveys. This testing can also be performed at scale for less effort relative to qualitative research with humans.

In the sections below, we'll outline what goes into creating these animated personas and how they are employed in research.

Why use TAPs in research?

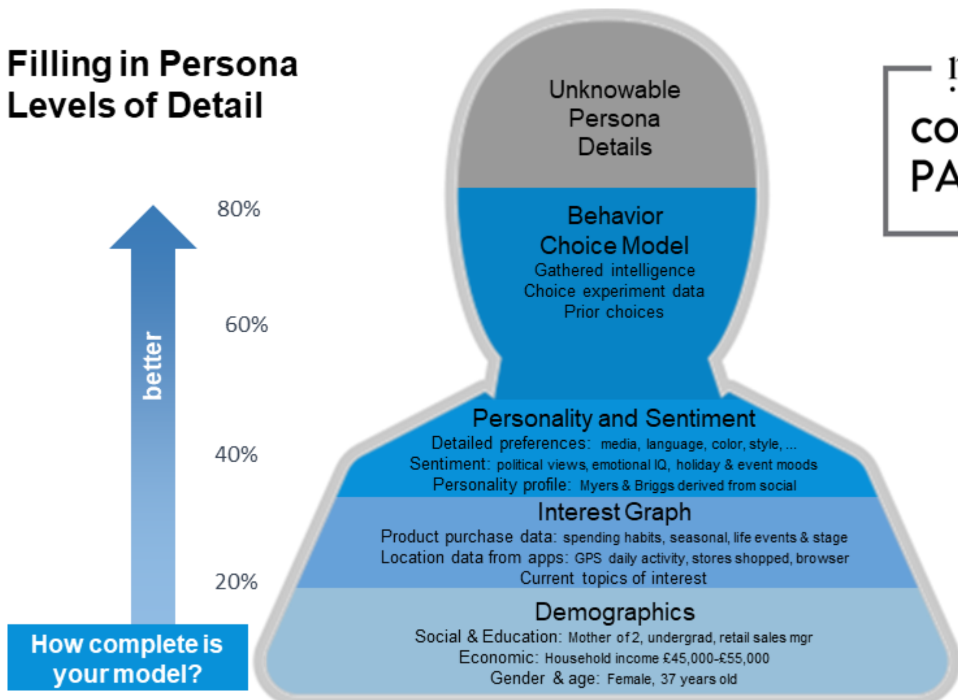
Modeling and Simulation

TAPs are essentially agent-based models that can be placed in simulated environments to see how their models will react to different stimuli. The models are created using any of three methods:

1) Derived completely from data
2) Derived partly from data and partly hand constructed using a set of assumptions
3) Created completely by hand from assumptive models.

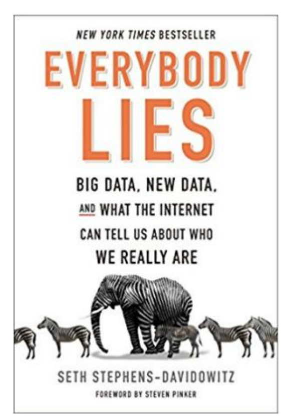
Each of these methods can be useful for testing ideas and predicting outcomes.

We typically seek to create a model of a persona from data such as purchase patterns or social media interaction, then augment that model with assumptions about other possibly unknown variations such as geography, political beliefs and other psychographic models to generate subtypes of a persona (for example, Soccer Mom at 25 living in Jacksonville vs. Soccer Mom at 43 living in Seattle). The level of detail of the persona, and therefore the nuance of responses to stimuli, increases in relation to the diversity of data used to construct it. In the simulated TAP environment, brands can run tests to see how each subtype of persona might react to new product offerings or messages, allowing for both unprecedented scale and granularity in market research.



nielsen
CONNECTED PARTNER

The Problem with Focus Groups, Surveys and ePanels

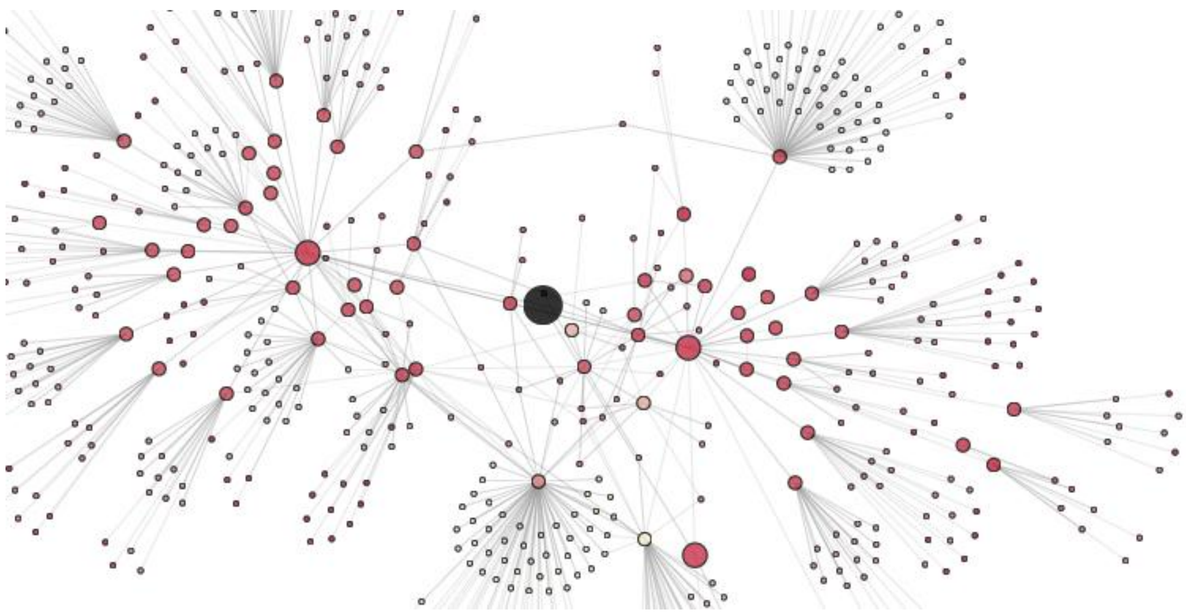


When Netflix sought to improve their recommendation system they issued a $1 million grand challenge and had people around the world compete to improve their recommendations. The winning team improved on Netflix's own algorithms by only 10%. Prior to the challenge, however, they had determined that asking people a set of survey questions yielded very different recommendation results from actual user activity including ratings. **In other words, there is a marked difference between what people say and what they do.** The best indication of a person's real interests is what they do with their attention. Therefore, TAP models created from strong historical data can actually be superior in their predictions to models based on questionnaires. For example, past buying behavior is a better predictor of future buying behavior than surveys will reveal.

How we create a TAP

Step 1: Gather as much data as we can on the target population: demographic, psychographic, sentiment (writing samples or surveys?), and choice model data from either a discreet choice model experiment on the target population or purchase and watch data.

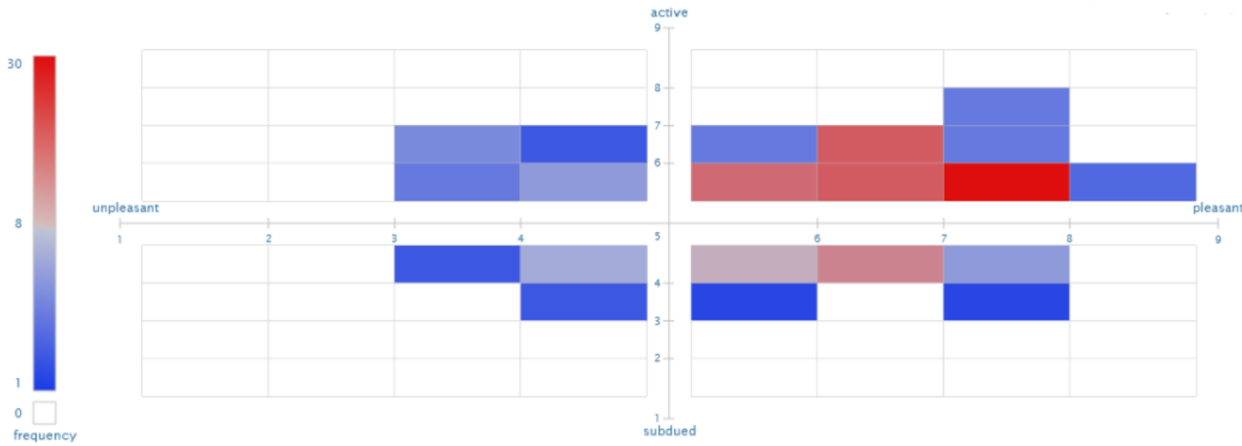
Step 2: The machine learning system, along with a human analyst, generates a list of topics, areas of interest, specific interests, sentiments and choice preferences. Those are correlated and clustered into pattern groups. We can also create them on specific historic people given enough data: For example, Martin Luther King Jr's body of writings is fed into Tanjo, which determines what conceptual vectors and topics make up MLK's thinking. This includes not only predictable concepts such as civil rights, but also unexpected concepts that only the machine detects. This map of interests and concepts and sentiments then makes up the persona model.



An interest graph is what makes up the "brain" of the persona. It can be visualized like the photo above, with hierarchies of interests, and concepts that make up the larger whole.

Interest graph with hierarchical interest area nodes and corresponding weights (strengths) for each

Step 3: If we have sample writing or other social listening data we can also build deep psychographic models into the personas and they will demonstrate preferences for certain language in written (and later spoken) conversations.



Beyond just what a persona is interested in, psychographic models can determine what sentiment they prefer in a given text (visualized above). For example, a Bernie Sanders persona would be very interested in an article about Trump, but not because he likes Trump.

Step 4: Once configured and activated on the Tanjo server, the persona model can be exposed to online content, articles and videos for information objects that match their interests and preferences. They will read those items or "Watch" the videos (currently they read the closed captioning text of a video), and rank the content they encounter with an interest score, building an interest graph.

> **Note:** There are 2 categories of interest graph, (1) the sub-topic interest graphs for each persona and (2) the higher order interest graph that aggregates both the interests and content from sub-topics. This is why you get a report of two scores for each message:
> (1) an overall score that rolls up all of a persona's interests, and
> (2) the score of the one sub-topic that evoked the highest interest.

Step 5: The scanning and scoring in Step 4 repeats for the number of topics defined in Step 2.

Step 6: After all topics have been scored, calculate the word cloud.

At this point, you have a persona with a complete hierarchy of interest graphs. The Tanjo system continues to scrape tens of thousands of sources every day to pull in the most popular and relevant content. Each piece of content is held up, and scored against the interest graphs of the personas, and the persona's topic maps.

In this way, the personas "experience" new information and current events and – if desired – their attitude interests can evolve over time. One of Tanjo's earliest personas, a simulation of Victor Hugo, was created in February of 2017. We've observed that, based on the content read over the past year, the Victor Hugo persona has diverged in its interest graph and preferences from the persona that was initially created. This indicates that he has been influenced by the media to which he has been exposed, which is of philosophical interest, but may not match a client's research goals. Typically our consumer marketing clients prefer that the personas not change, so we have the ability to lock the interest graph when the persona is initially created.

Research Applications

Here are two key benefits to marketers and market researchers.

Benefit One: Test messaging

Beyond just watching the personas to see how their interest graph changes as they react to current events, Tanjo's toolset provides the ability to present new writing to the personas and have them respond with their interest scores. The personas will react in real-time with interest scores from their individual perspectives. Testing before sending emails or launching ad campaigns – and tuning messaging for each segment – can dramatically improve response.

Benefit Two: Circumvent marketing and sales bias

A very illustrative and useful exercise is to build persona models based on a survey of the marketing personnel to reveal who they think their different customer segmentation models are; then create persona models from data or customer segmentation models and compare the two to see where there are differences between perception, and what the data reveals. This typically reveals new insights that indicate new approaches to messaging.

Deep Dive: Behavior Trees

What follows is a description for TAP behaviors with completed choice models defined in hierarchical behavior trees. (improving the model to 80%) This method was developed during our tenure at Lockheed Martin building large agent-based model simulations.

The basic framework of this architecture is a modular, hierarchical decision making approach, similar to the popular Behavior Tree (BT) architecture used in games, called the Component Reasoner. It can support many approaches to decision making, but we rely primarily on a utility-based approach called the Weight-Based Reasoner created at Lockheed.

BTs have two major advantages. First, the hierarchical approach is extremely powerful, avoiding spending processing time on irrelevant decisions, and is a natural way to structure the AI such that independent decisions are decoupled. Second, the options are modular, which is to say that a given option can appear multiple places in the tree. This prevents the need to re-implement functionality every place that it is used.

By design, BTs rely on simple Boolean discriminators for their selectors. This simplifies implementation, but puts a limit on how well the AI can examine the subtle nuance of a situation before making a decision. More generally, it has been our experience that there are often cases where a more complex approach to decision making should be used for a particular decision, while retaining simplicity elsewhere.

Thus we want a framework which retains the hierarchy and modularity of the BT's structure, but allows us to employ complex decision makers where appropriate while retaining support for simple selectors elsewhere. The Component Reasoner does this by using *reasoners* rather than selectors to make each decision.

The difference between a reasoner and a selector is subtle but important. Selectors are expected to use simple logic, such as taking the first valid option, selecting each member of a sequence in order, or using fixed probabilities (assigned a priori, not at runtime) to make their decision. In contrast, a reasoner is allowed to be arbitrarily complex.

The advantage of this approach is that it allows us to select the most appropriate approach for each decision being made. For decisions that are highly deterministic we can use a BT-style selector. For decisions that require us to weigh the relative advantages of several possibilities, a utility-based approach will work well. In a situation where the AI needs to learn from past results, we might attempt something like Genetics-Based Machine Learning (Harrison 2007).

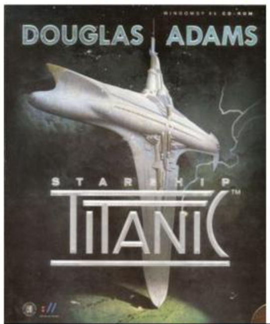
The root of the Component Reasoner contains a single *option*. An option is a structure which contains one or more *actions*, all of which will be executed if that option is selected (the root option is always selected). As in a BT, these actions may be *concrete* or they can contain another reasoner. If they contain a reasoner, it can use whatever approach makes the most sense for that particular decision to pick from among its own options. Control works its way from the root option down through its actions to the reasoners they contain, into the options those reasoners select, and so on until we reach the concrete actions in the leaves.

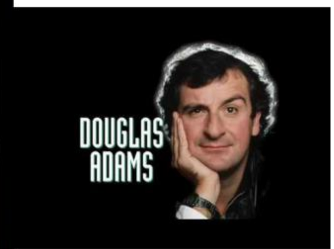
It's worth emphasizing that this structure supports parallel execution. Because an option can contain multiple actions, including multiple sub-reasoners, it can do more than one thing – and even make more than one set of decisions – at once. This is a capability which is missing from all too many AI techniques.

Future work: Natural language discussions

In 1997 members of the Tanjo team worked on a game with science fiction writer Douglas Adams (Hitchhiker's guide to the Galaxy) called *Starship Titanic*. In that game we attempted to create a means to converse with characters in natural language. Our entire Velocitext dictionary contained about 500 words, and yet some of the resulting interactions during gameplay caused Douglas to coin the term "Spookitalk" to describe the eerie feeling of interacting with a synthetic character. Today we have the means to go far beyond those early attempts and expect to permit TAP users to speak freely and convincingly with their customer persona models in the future.

1997 Starship Titanic NLP





In 1997 Douglas Adams tapped us to crea
natural language processing system to er
conversation with the game characters ir
"Starship Titanic". He ended up calling it
Spookitalk to describe the feeling on the
the neck we sometimes got when the ga
characters responded in a surprising way

Interesting how this scene looks like the robot bartender in the Chris Pratt/ Jennifer Lawrence film "Passengers"

BILLARI, F. C & Prskawetz, A. (2003) *Agent-based computational demography: using simulation to improve our understanding of demographic behaviour*. Springer, Heidelberg: Physica Verlag.

BILLARI, F. C., Prskawetz, A., Diaz, B. A., & Fent, T. (2008). The "Wedding-Ring": An agent-based marriage model based on social interaction. *Demographic Research*, Volume 17, 17, 59. [doi://dx.doi.org/10.4054/DemRes.2007.17.3]

BILLARI, F.C. (2006). Bridging the gap between micro-demography and macro-demography. In Caselli, G., Vallin, J., and Wunsch, G. (eds.), *Demography: analysis and synthesis*, Vol 4. New York: Academic Press/Elsevier: 695–707.

DEAN, J. S., Gumerman, G. J., Epstein, J. M., Axtell, R. L., Swedlund, A. C., Parker, M. T., & McCarroll, S. (2000). Understanding Anasazi culture change through agent-based modeling. In Kohler, T. A. & Gumerman, G. J. (eds) *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes*, New York: Oxford University Press, 179–205.

DIAZ, B. A., & Fent, T. (2006). An agent-based simulation model of age-at-marriage norms. In Billari, F. C., Fent, T.. Prskawetz, A. & Scheffran J. (eds.), *Agent-Based Computational Modeling. Applications in Demography, Social, Economic and Environmental Sciences*, Heidelberg: Physica-Verlag, pp. 85–116.

EDMONDS, B., & Moss, S. (2005). From KISS to KIDS–an 'anti-simplistic' modeling approach. In Davidsson, P., Logan, B. & Takadama, K. (eds.), *Multi-Agent and Multi-Agent-Based Simulation*, Heidelberg: Springer Verlag, pp. 130–144.

FLACHE, A., Macy, M. W., (2005). Social life from the bottom up: Agent modeling and the new sociology. In C. M. Macal, D. Sallach, & J. North (Eds.), *Proceedings of the Agent 2004 Conference* (No. 2004). The University Argonne National Laboratory of Chicago.

HEDSTRÖM, P. (2005). *Dissecting the social: On the principles of analytical sociology* (pp. 114–44). Cambridge: Cambridge University Press.

LANSING, J. S. (2002). "Artificial societies" and the social sciences. *Artificial Life*, 8(3), 279–292. [doi://dx.doi.org/10.1162/106454602320991864]

MACY, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual review*

of sociology, 143–
166. [doi://dx.doi.org/10.1146/annurev.soc.28.110601.141117]

MOSS, S., & Edmonds, B. (2005). Towards good social science. *Journal of Artificial Societies and Social Simulation*, 8(4), 13. http://jasss.soc.surrey.ac.uk/8/4/13.html

SILVERMAN, E., Bijak, J., & Noble, J. (2011). Feeding the beast: can computational demographic models free us from the tyranny of data? In: T. Lenaerts, M.Giacobini, H. Bersini, P. Bourgine, M. Dorigo and R. Doursat (eds.) Advances in Arti cial Life, ECAL 2011, MIT Press, Cambridge MA, USA (pp. 747–
754). http://eprints.soton.ac.uk/272839/. Achieved at: http://www.webcitation.org/6dQ9GYpxl

SIMÃO, J., & Todd, P. M. (2003). Emergent patterns of mate choice in human populations. *Artificial Life*, 9(4), 403–
417. http://www.ncbi.nlm.nih.gov/pubmed/14761259. [doi://dx.doi.org/10.1162/106454603322694843]

STERMAN, J. D. (1991). A skeptic's guide to computer models. Managing a nation: The microcomputer software catalog, 2, 209–
229. http://www.systems-thinking.org/simulation/skeptics.pdf. Achieved at: http://www.webcitation.org/6dQ9q5Ot2

TODD, P. M., & Billari, F. C. (2003). Population-wide marriage patterns produced by individual mate-search heuristics. In Billary, F. C & Prskawetz, A (eds), *Agent-based computational demography: using simulation to improve our understanding of demographic behaviour*, Heidelberg: Springer Physica Verlag, pp. 117–
137. [doi://dx.doi.org/10.1007/978-3-7908-2715-6_7]

WINKLEBY, M. A., Jatulis, D. E., Frank, E., & Fortmann, S. P. (1992). Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *American Journal of Public Health*, 82(6), 816–820. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1694190/. [doi://dx.doi.org/10.2105/AJPH.82.6.816]

WU, M. B. (2012). A hybrid microsimulation model for a UK city population with dynamic, spatial and agent based features. University of Leeds. http://etheses.whiterose.ac.uk/id/eprint/3758. Achieved at http://www.webcitation.org/6dWBwsDQv

The rest of this page intentionally left blank