Intrinsic Elicitation: A Model and Design Approach for Games Collecting Human Subject Data

David Gundry
University of York
York, UK
deg500@york.ac.uk

Sebastian Deterding
University of York
York, UK
sebastian@codingconduct.cc

ABSTRACT

Applied games are increasingly used to collect human subject data such as people's performance or attitudes. Games afford a motive for data provision that poses a validity threat at the same time: as players enjoy winning the game, they are motivated to provide dishonest data if this holds a strategic in-game advantage. Current work on data collection game design doesn't address this issue. We therefore propose a theoretical model of why people provide certain data in games, the *Rational Game User Model*. We derive a design approach for human subject data collection games that we call *Intrinsic Elicitation*: data collection should be integrated into the game's mechanics such that honest responding is the necessary, strategically optimal, and least effortful way to pursue the game's goal. We illustrate the value of our approach with a sample analysis of the data collection game *Urbanology*.

CCS CONCEPTS

• **Applied computing → Computer games**;

KEYWORDS

Applied Games, Games with a Purpose, Human Computation Games, Crowdsourcing Games, Human Subject Data, Intrinsic Integration, Validity

ACM Reference Format:
David Gundry and Sebastian Deterding. 2018. Intrinsic Elicitation: A Model and Design Approach for Games Collecting Human Subject Data. In *Foundations of Digital Games 2018 (FDG18), August 7–10, 2018, Malmö, Sweden.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3235765.3235803

1 INTRODUCTION

Games and game design are increasingly popular ways to elicit data from people, particularly large online populations [2, 25, 63, 74]. Such applied games for data collection fulfil a dual function: they (1) structure a data-providing task and (2) motivate participation by turning data provision into enjoyable gameplay. For example, the *ESP Game* turns the tagging of images into a game of mind-reading: players are randomly paired with other anonymous players, presented with a series of images, and for each, have to guess and type what word the other player would associate with it [1]. Thus, the game (1) structures the task (providing concrete instructions, goals, materials, tools), and (2) motivates it, evidenced by 13,000+ players voluntarily producing 1.2 million labels in the course of four months [1]. A key factor of the success of such data collection games is their design [2, 19, 61]. Existing work has explored design strategies for *either* motivating players *or* ensuring data quality after collection [56, 60]. Validation strategies for the latter include agreement designs, reputation systems, and automatic checks, while motivation strategies have focused individual game design elements like rewards, effectively 'gamifying' data collection games. This overall gamification+validation approach frames the key design challenge of data collection games as one of post-hoc filtering and checking a maximized volume of player-provided data. As such, it treats player motivation and data quality as *separate concerns*. Yet different forms of motivation have been shown to directly affect the kind and quality of data provided [62, 82]. This marks a significant shortcoming of gamification+validation approaches and invites the search for broader *elicitation approaches* that integrate *both* motivation *and* data quality: models and design principles for motivating players to provide desired data in a desired quality from the outset. Such approaches are of particular relevance where there is no straightforward way of post-hoc validating data against an objective or consensus ground truth, e.g. when eliciting human subject data such as people's preferences, beliefs, or performance.

In this paper, we introduce such an integrated design approach which we call *intrinsic elicitation*, akin to the principle of *intrinsic integration* in e.g. educational games or gamification [23, 41]. In short, intrinsic elicitation captures the idea that generating desired data in a desired quality should be integrated into the mechanics of the game in such a way that it is the necessary, strategically optimal, and least effortful way for the player to pursue the game's goal. We will develop and defend this approach as follows: First, we situate data collection games within the wider field of applied gaming. We argue that data collection games *as games* introduce a systematic motivational driver *and* threat to data validity at once which existing gamification+validation work hasn't addressed. Therefore, we present a theoretical model of why players provide particular kinds of data in a game, the *Rational Game User Model*, integrating Jonas Heide Smith's *Rational Player Model* [75] with existing theoretical and empirical work on data collection games and survey engagement. From this model, third, we derive the design approach of intrinsic elicitation, comprising three heuristics for how to integrate data collection into a game's mechanics: necessity, centrality,

FDG'18, August 7-10, 2018, Malmö, Sweden

ACM ISBN 978-1-4503-6571-0/18/08…$15.00
https://doi.org/10.1145/3235765.3235803

and veracity. We illustrate the utility of our approach as a predictive model and evaluative tool by analysing the location-based data collection game *Urbanlogy* [15] through its lens.

2 BACKGROUND

2.1 Applied Games for Data Collection

Applied gaming is the design and use of games to achieve an ulterior purpose [66], such as learning in game-based learning, belief and attitude change in persuasive games, or changed health beliefs, attitudes, and behaviours in games for health. One growing domain of applied gaming is research and data collection, sometimes called "gamifying research" [25] or "game-based methods" [74]. Examples abound, with games prompting players to discover protein and RNA structures (*Foldit*, *EteRNA*), report and classify bird sightings (*eBird*), classify images of galaxies *Galaxy Zoo* [17], reveal their cognitive processes (*The Great Brain Experiment* [10]) or assess their fluid intelligence [31]. Outside science, we find data collection games for digitizing and labelling archive material [16, 30], tagging images[2], or codifying knowledge for machine learning [3]. The literature uses several terms to denote these kinds of games [67]: "citizen science game" [17] foregrounds the participation of citizen volunteers in science. "Crowdsourcing games" [39] denotes the enlistment of large online populations in a task. "Human computation games" and "games with a purpose"[2] revolve around the offloading of computing tasks to humans. Related to but distinct from such data collection *games* is *gamified* data collection – using game design elements [23] to e.g. motivate participation in surveys [48] or crowdsourcing [54]. Finally, *game intelligence* describes the opportunistic (re)use of entertainment games data for scientific purposes [27].

No matter the label, the main reason for using game-based data collection has been participant motivation. Applied games harness the appeal of games for their ulterior purpose, hoping to attract and motivate large populations to play voluntarily, thus increasing participation and retention rates and reducing the need for costlier means of motivation such as payment [17, 63]. Applied games therefore have a dual, interlinked design goal [23]: they need to (a) motivate players to (b) engage in play that produces the desired outcome – valid data in the case of data collection games. With'valid', we here mean more generally data of a kind and quality that is fit for its intended purpose, which may be drawing valid inferences (standard scientific validity [53]), making good hiring decisions based on in-game performance, directing crisis responders to likely survivors spotted on satellite imagery of a flooded region, etc. No matter the purpose, if gameplay is engaging but produces little or invalid data, the game is ineffective. Similarly, if gameplay generates valid data but is not enjoyable, players won't come and generate the desired quantity of data.

2.2 Current Design Approaches and Research

For these reasons, numerous researchers have studied how to design data collection games for motivation and data quality. Several studies have probed the underlying motives of players, converging on constructs of *gaming enjoyment* (e.g. fun, competence, relaxation), *community participation* (status, recognition, social norms), and *meaning* (making a contribution to science, helping others, growing oneself), with *monetary benefits* being rarely used and reported [20, 37, 38, 78]. In their systematic analysis, Tinati and colleagues [78] suggest that these motives can be classified following self-determination theory [21] into *intrinsic motivations* (gaming enjoyment factors like competence, the community factor relatedness, and all meaning factors) and *extrinsic motivations* (monetary benefits, achievement, status, recognition, social norms). This is particularly relevant as research on crowdsourcing and survey design suggests that intrinsic motivation leads to higher participation rates *and higher-quality data* [62, 80, 82] – comparable work on data collection games is unfortunately harder to find [58, 59].

As for *motivational design strategies*, the literature has chiefly explored individual game design features such as difficulty balancing [64], visual appeal [81], graded goals [34], narrative and theming [59], and reward systems [33, 39, 59, 72], though with various results. In a sense, the existing literature has been less concerned with what core loops and mechanics [23, 69, 71] fit what data collection tasks than with 'gamifying' data collection games – adding and tweaking presumed-engaging design features. This is arguably because the majority of data collection games follows the template of the early successful and much-publicized *GWAPs* of Louis von Ahn and colleagues[2]: players classify or transcribe presented (usually visual) data, receiving points for every (correct) input. A second such influential template is *Foldit* by Seth Cooper and colleagues [17], where players generate solutions to problems where the theoretically possible optimum is known but not the actual best possible solution; here, players score based on the number and optimality of provided solutions. There are good practical reasons for the popularity of these templates: they offer working models that are easily replicated via open access platforms like *Galaxy Zoo*; they address classification and solution discovery tasks with broad applications; and they provide straightforward means of *validating generated data* – the second main design concern of data collection games.

Commonly used *validation strategies* are agreement designs, automatic solution evaluation, and reputation systems [19]. In agreement designs found in *GWAPs* and most data classification games, a player's input is assessed on how much it agrees with the inputs of other players on the same stimulus or task [2]. Poor responses are filtered out or demoted in their weighing as they are unlikely to 'agree' with the consensus of the majority and/or trusted players. This is somewhat data-inefficient as it requires multiple people to solve the same task. In contrast, solution discovery games like *FoldIt* [18], automatically evaluate the quality of each submitted solution against known and computationally formalized optimality criteria. Solutions are ranked and scored based on how close they come to the theoretical optimum. While potentially more data-efficient than agreement designs, this validation strategy obviously requires prior knowledge of solution requirements that can be computationally expressed and validated. Finally, both agreement and automatic evaluation designs often feed into reputation systems that track player performance over time to identify which players reliably provide high- or low-quality data [19]. These reputation scores can then be used to weigh answers in agreement designs, filter out data by low-scored players, or optimize player-task matching, e.g. serving difficult or unsolved tasks to high-scoring players first [64].

To summarize, current design research on data collection games has chiefly focused on variations of *GWAP*-style data classification games and *Foldit*-style solution discovery games, replicating their core game mechanics based on validation strategies [17, 57, 60]. In this, research has treated player motivation and data quality as mostly separate design concerns addressed with separate solutions: 'gamifying' data generation with reward systems etc. to maximise data *volume*, then validating generated data with agreement, automatic evaluation, or reputation systems to maximise data *quality*. This gamification+validation approach works with the somewhat wasteful assumption that some or even significant amounts of poor-quality data are inescapable: as long as *some* players provide good data, ground truth will out and can be used to identify and reward high-quality data. More importantly, this approach necessarily requires *some* validation against a known objective or consensus ground truth.[1]

2.3 The Challenge of Human Subject Data

Exactly this requirement sets *GWAP*- and *Foldit*-like games apart from games designed to collect *human subject data* like short-term memory processes [10] or people's performance [6]. It also limits the applicability of their underlying gamification+validation approach.[2] Contrary to classification tasks or solution discoveries, for human subject data, the ground truth is often unknown and *unknowable* to anyone but the subject, and data validity can not be equated with players doing 'as best they can'. For subjective attitudes, values, or preferences, there is by definition no subject-external ground truth to assess them against. We often aggregate such data ('people on average give this service a 7.5 net promoter score'), but in that usually want each subject to honestly report their independent evaluation, not their 'best guess' at what an average evaluation would be. Similarly, much of human subject research is interested in covert, non-conscious processes, tendencies, dispositions, states, or traits that reveal themselves in people's 'spontaneous' responses, e.g. the preferred walking speed as an expression of fitness or wage levels and derived valuations of time [11, 50]. In these instances, the moment one communicates one answer to be 'more true' or 'more optimal', this would distort the generated data. Even where there are operationalizable scales for performance (such as IQ or money earned in a game theoretical experiment), people may be motivated to overstate (or understate) their 'true' performance ability because it is socially desirable or rewarded by the game. And again, for individual capabilities like IQ, there is no subject-external ground truth to assess how accurately the subject's current recorded performance reflects its 'true' underlying capability.[3]

More generally, the collection of human subject data introduces specific data types, validity criteria and validity threats that gamification+validation approaches don't reliably address. Worse, where gamification+validation approaches model or even reward certain responses as 'better' or 'more true' than others, they generate particular new validity threats. In education and gamification, these threats have been discussed as *gaming the system* [4, 7, 83] or *cheating* [52]. Individuals game the system when they find ways within the rules of a system to maximise their evaluation metrics at the expense of the substantive goals intended by the system, e.g. giving short nonsense answers in a question and answer platform because every answer receives points irrespective of content. Individuals cheat when they covertly gain an advantage through means outside the rules of a system, such as secretly copying answers for a test from a colleague.

This raises the obvious question what game design approaches are better suited to human subject data than the currently prevalent gamification+validation approach. Education and gamification research are plausible sources of alternatives since both often involve human subject data collection and have dealt with gaming the system and cheating. One consistent argument across these two fields, going back to Thomas Malone [51], is that the outcome of applied gaming – teaching particular skills, making a particular activity more engaging – should in some way be *integrated* into the game's mechanics [23, 41, 77]. In game-based learning, this principle is called *intrinsic integration* [41]. It is 'intrinsic' in that (a) the learning material is part and parcel of the enjoyable, intrinsically motivating core mechanic of the game, and (b) the game's mechanics and thematic world embody and represent the learning material. Deterding [23] suggests in direct analogy that effective gamification is intrinsically integrated by turning the target activity into the core mechanic of the game, reorganizing its 'core loop' to support intrinsic motives like competence. There is some evidence that intrinsically integrated educational games and gamified interventions outperform their alternatives [12, 28].

This notion of intrinsic integration is not without parallels in data collection games. Tuite [79] cautions that GWAPs should "match mechanics to purpose". Observing that existing validation templates are insufficient to design new GWAPs, Galli [32] offers a range of standard game mechanics that match different GWAP task types, e.g. memorisation maps clustering. Jamieson, Hall and Grace [44] suggest identifying mechanics for human computation games by finding mechanics or real-world activities that are "isomorphic" to the structure of the computational task. However, Jamieson and others argue for this as a way to reduce extraneous effort and cognitive load and ease problem-solving, and none of them address the particular validity threats of games for collecting human subject data. For these, we need an intrinsic integration approach that not only matches a mechanic to a particular data type, but also *motivates* players to provide honest data without biasing results, especially in instances where there is no subject-external ground truth to validate responses against. Such an approach arguably requires developing a clear idea of what motivates players of data collection games to take particular in-game actions (and thus provide particular data) rather than others. That is the purpose of the next section.

3 THEORY

3.1 The Rational Player Model

Implicitly or explicitly, game designers, researchers, and members of the public hold different mental models of players [75]. For instance, we may view players as mostly passive objects of game stimuli (a view proposed by strong media effects research), or as highly

[1] See e.g. Siu, Zook and Riedl's [73] framework of mechanics for human computation games, which includes validation as a necessary component.

[2] For clarity, we here delineate *human subject data collection games* as games intentionally designed and used to generate data about the playing human participants through play, e.g. to assess their performance, survey their attitudes and preferences, or gather data for basic and applied human and behavioral science research.

[3] Beyond these instrumental challenges, human subject data collection games open a plethora of ethical questions, especially if assessment occurs covertly under a playful veneer. For reasons of space, we cannot engage with these here, though see [26] .

autonomous subjects appropriating games as a mere material for their own ends, as in e.g. Sicart's humanist rhetoric of play [70]. The model underlying much if not most game design practice is what Jonas Heide Smith has elucidated as the *Rational Player Model* [75]. In short, the Rational Player Model states that in-game, players are self-determined and rational actors "whose main (or only) concern is to optimize [their] chances of achieving the [game's] goals." [75] Put more simply, people play to win.

This model holds substantial merit despite and because of its simplicity. Games are often distinguished from toys solely through the possession of a goal [46]. Striving towards goals is widely seen as what distinguishes 'gaming' from 'playing' [23]. Across studies of player motivation and experience, players state that experiences of competence, mastery, and achievement gained from achieving game goals make playing engaging [8]. In addition, 'playing to win' is an important and actively sanctioned social norm of most gaming encounters: to make no visible effort to win during gameplay usually results in being reprimanded as a'spoilsport' (see[22], pp. 174-5 for a review). Gaming is one of the few types of social situations where "setting aside all personal feelings and all impulsive inclinations" to rationally maximise one's own goal attainment is allowed and indeed expected ([36], p. 96). And in any domain of everyday life, goals are extensively and effectively used to motivate and direct effort [40]. Within a game, goals have a similar function, directing player effort towards particular future states.

Starting with the assumption that players try to act strategically optimal to attain a game's goals is not only well-supported: it also opens the way to powerful conceptual tools for analysing and predicting how particular game design choices will affect in-game actions, as Smith demonstrates in his formalisation of the Rational Player Model [75]. He explicates the Rational Player Model using *game theory*, the mathematically formalized study of *strategic interaction* when two or more actors make decisions with clearly defined objectives, taking their knowledge and expectations of the *other* actors' objectives and decisions into account [55]. As such, game theory shares basic assumptions (and mathematical inclinations) with rational choice theory in sociology and economics, namely that people are individual agents acting to rationally maximize their personal utility [68]. Translated into gameplay, players are rational agents seeking to optimise their utility as defined by the game's goals. In this view, 'gaming the system' as defined previously is normal and indeed *expected* gaming behaviour, as is cheating: cheating is likely when the expected utility of cheating (minus the expected disutility of the chance of being caught) is bigger than the utility of available alternative actions.

Being caught leads us to the point that player's in-game choices are affected by more considerations than winning. Juul [47] for instance articulates at least three concerns from which players can and often do assess in-game actions: The first is goal-orientation or the desire to win, matching the Rational Player Model. This is nested in a concern for gameplay as an interesting experience: players try to maximise their enjoyment, perhaps by playing in a way that is strategically sub-optimal but more novel and interesting. Even this is coached in a third wider consideration of the social implications of game actions. We often self-handicap when playing against children, for example. However, all these concerns do not speak against a game theoretic analysis of gameplay. Rather, they

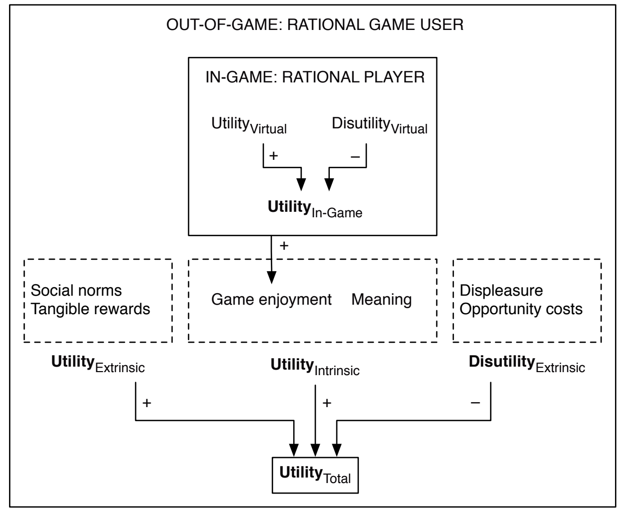

Figure 1: The Rational Game User Model.

highlight a common narrow misconstrual of the concept of *utility*. Originating in Bentham's utilitarianism, utility expresses the tendency of an object or action to "produce benefit, advantage, pleasure, good, or happiness" (and reduce or prevent the opposite) [9]. In later economics, this substantive conception has been replaced by an operational definition of utility as the *preferences* of individuals, as revealed in their choices [9]. Either conception is perfectly in keeping with any and all considerations Juul (and others) have brought forward as informing in-game actions. If we value and derive pleasure from making a child happy, and more so than winning ourselves, then self-handicapping is the rational choice that maximizes our utility, as revealed in our choice to self-handicap. The utility that a rational player seeks to maximise can include the joy of winning, other joys of gameplay, and surrounding social norms at once.

3.2 The Rational Game User Model

For the design of data collection games, we therefore suggest a strong rational choice-style abstraction that starts with the strategic utility of in-game actions for accomplishing game goals. We see five main benefits to doing so. Firstly, this 'in-game' utility elegantly compresses many known important considerations of players outlined above, from mastery to the social norm of 'playing to win'. Second, Smith's own empirical work indicates that players do rationally play to win in-game and modulate arising social concerns through parallel out-of-game talk[75]. Third, it is literal textbook practice and therefore easy to adopt by practitioners: game design textbooks regularly advise to use game theory to calculate the strategic utility of in-game actions and objects as part of balancing to afford interesting decisions and a perceived 'fair' chance at winning [65]. Fourth, it allows to articulate clear, mathematically expressed hypotheses and predictions, enabling rigorous empirical testing and robust design guidance. Fifth and finally, we concur with Healy [42] that 'nuance' in theory is overrated: theory is more pragmatically powerful when it needs *less* starting assumptions and data to make good-enough predictions, and knowledge generation can be more usefully guided and integrated by extending simple models with variables like 'social context' only if and when the data requires it.

Still, we recognise that in-game actions do not occur in a vacuum. For human subject data collection games in particular, we need to consider what utilities existing research has identified beyond game enjoyment. Here we can draw on literatures on motivations for participating in human subject research (particularly online surveys), volunteer crowdsourcing, and citizen science, as they are structurally comparable to data collection games. For web surveys in specific, researchers have developed and tested rational choice models that predict participation decisions [29]. Surveying major recent reviews [5, 29, 35, 62, 80], a picture broadly concordant with the data collection game literature emerges, notably excluding gaming enjoyment factors. First and foremost, participation is intrinsically motivated by *meaning* factors: helping others, contributing to science, growing oneself through learning. Second, participation is extrinsically motivated by perceived social norms like reciprocity and, where offered, tangible rewards like money. Third and finally, participants take into account the *disutility* of labor involved in participating, such as the opportunity costs of foregone alternative ways of spending the same time and the active displeasure of doing something boring or strenuous [76]. In online surveys for instance, overly long surveys or poor usability lead to low participation, high abandonment, careless responding or satisficing: doing just enough to achieve a somewhat satisfactory answer [29, 49].

Given the general concordance of the different motivation literatures, we assume that these out-of-game factors or utilities also operate when players choose which in-game action to take (and thus data to provide) in a data collection game. We summarize these factors in what we call the Rational Game User Model (illustrated in figure 1): As a rational game user, players want to maximise their **total utility**. This includes out-of-game **extrinsic utilities** like *social norms* and *tangible rewards*: playing the game because doing so is incentivized and/or socially expected or sanctioned. On their own, extrinsic utilities will motivate players to play the game/provide data *just enough* to satisfy incentive criteria or social expectations with minimum effort, honesty, and care: they invite gaming the system and careless responding [62, 80, 82]. They also invite dishonest responding [45] if participants consider it socially desirable to over- or under-report certain traits or beliefs. This stands in stark contrast to **intrinsic utilities**, especially *meaning*. As players are motivated to help others or science and find the game itself a valid means of doing so, they will attempt to provide data that optimally serves the game's ulterior purpose, e.g. answering an open-ended question in detail and truthfully, despite the involved effort or social undesirability of the answer.

The second, game-specific intrinsic utility is *game enjoyment*. This is where the model incorporates the rational player. To maximize game enjoyment, the rational game user analyzes the current game state as a rational player trying to maximize their **in-game utility** as expressed in the game's goals. To do so, they assess each currently possible action for its *virtual utility* (how much or likely the action moves them closer to goals and *virtual disutility* (the opportunity costs of in-game resources spent on the action). While game enjoyment uniquely motivates data provision in data collection games, it also opens a second, game-specific route to the validity threat of gaming the system: if an action/data provision maximizes in-game utility more than the more honest or 'spontaneous' option, players are systematically more likely to choose the more in-game optimal and thus enjoyable action.

Finally, the rational game user model acknowledges that data provision is effortful labor – else, there would be no need to motivate it with a game. Thus, beyond extrinsic and intrinsic utility, a rational game user will consider an action's **extrinsic disutility**. Playing a game involves effort and displaces other activities we could have done instead. One may object that because gameplay is intrinsically motivating and enjoyable, it lacks the common displeasures of labor and should be one of the most highly preferred activities. But players do regularly report negative experiences like frustration during gameplay e.g. due to poor playability and usability[38], and phenomena like goldfarming demonstrate that some aspects of gameplay (like 'grinding') have a high enough disutility that people pay money to free their time for other, more preferred activities[43]. Even in enjoyable games, players regularly satisfice, making good-enough choices instead of investing more time and effort into calculating the absolute optimal move. For data collection games, this means that all else being equal, responding honestly or spontaneously should be the most effortless option, or at least as effortless as any other available choice. As displeasure and opportunity costs increase, players will be more likely to respond carelessly, satisfice, or even abandon the game.

4 INTRINSIC ELICITATION

From the Rational Game User Model, we can derive requirements for translating human subject data collection into applied games, specifically their mechanics and core loops, which are commonly considered the primary formal aspects of a game[23]. *Game mechanics* refer to the methods by which an in-game agent effects a game state change [69]. They are the verbs of the game, like 'jumping', 'shooting,' or 'drawing a card'. *Loops* describe cycles of mechanic actuation, system processing, and system feedback relative to one or more game goals [23, 70]. Mechanics can be actuated in multiple ways with differing effort: we can jump high or low, and shoot with careful or careless aim. Thus, any in-game action involves two degrees of freedom: First, a player must choose *which* mechanic to actuate. Second, they choose *how* to actuate it. This latter *how* defines the sensitivity or expressive range of mechanics as measurement instruments. Just like a 5-point Likert scale can only support operationalisations that rely on the selection between five ordinal values, and endless runner game with a single jumping mechanic delimits measurement to timed button presses in response to on-screen events. Combining this with the Rational Game User Model, we derive three systematic principles for games that enable and motivate participants to generate valid human subject data. We summarisingly refer to these principles as the *Intrinsic Elicitation* approach to data collection game design 2:

(1) **Necessity:** Players only engage in data provision that changes game states. Requirement: Embody data provision tasks as the game's interesting mechanics.

(2) **Centrality:** Players select mechanics to maximise utility. Requirement: Make data-providing mechanics as strategically central and effortless as possible.
(3) **Veracity:** Players actuate mechanics to maximise utility. Requirement: Where honest responses are needed, ensure that they have the highest strategic utility and lowest effort, or at least the same utility and effort as any other available option.

In summary, Intrinsic Elicitation states that data generation should be integrated into the game's mechanics such that responding honestly is the necessary, strategically central, most in-game advantageous and least effortful choice for pursuing the game's goal.

4.1 Necessity

A rational player will only engage with game mechanics, because these are the only means to affect the game state and thus approach the game's goals. For the rational game user, any activity not related to mechanics only increases extrinsic disutility, unless it constitutes metagaming [70] like making the opponent nervous or satisfy extrinsic utilities like social norms. That is, the model acknowledges out-of-game actions, but suggests that designers can most reliably steer data provision through in-game mechanics. For data provision to occur, it must be instrumental in a mechanic. That is, it must be impossible to actuate the given mechanic without supplying the desired kind of data. Furthermore, the game mechanics themselves must be part of enjoyable game loops – actuating the mechanic in the pursuit of game goals should be an interesting challenge or decision that elicits experiences of curiosity, competence, achievement, and the like. If acting rationally to win isn't enjoyable, players are less likely to do it, or do it well [23].

Take *The ESP Game*[1] as an example. Two players each try to provide input that matches the other player's. Each round, the game provides both players with an image and prompts them to submit a written image label that they think the other player would use. The mechanic here is 'submit label,' which is actuated by typing one or more letters and hitting the submit button. Without doing so, and the game state doesn't change. Game mechanic and data provision are one and the same, and mind-reading others is an inherently interesting challenge. This necessity principle restates earlier suggestions that data collection games should "match mechanics to purpose" [32, 44, 79]. E.g., a game for assessing fluid intelligence should involve mechanics and goals whose successful accomplishment requires fluid intelligence, like *Portal 2* puzzles[31]. A game eliciting people's preferences in vacations should involve mechanics whose actuation expresses preferences, e.g. ranking photos of vacation places.

A corollary is that if data can be provided in different kinds that require different levels of effort, actuating the mechanic should *at minimum* require data of the target kind, not data of a kind that requires less effort. E.g., if we want to collect data about people's pronunciation of words with a game where they steer a plane by speaking, the game needs to be able to recognise and require actual spoken words. The mechanic should not be actuated by e.g. volume or pitch alone, as producing humming and nonsense sounds of different volume and pitch is likely less effortful than thinking of and speaking a large variety of actual words.

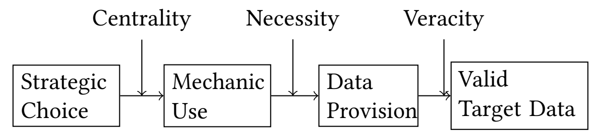

Figure 2: Intrinsic Elicitation

4.2 Centrality

Often the data-providing mechanic is not the only mechanic available in the game. This may be because the game designer wants to offer meaningful choice between different courses of action, interesting variety, or because the data-providing mechanic and surrounding loop are not inherently enjoyable and therefore require additional loops and mechanics that are enjoyable. In any case, the rational player will at any step select to actuate the mechanic with the greatest current perceived in-game utility. Therefore, the data-providing mechanic should be *central* to gameplay, either because it is the only or core mechanic the player necessarily actuates over and over (submitting words in *The ESP Game*, running in *Super Mario Bros.*), or because it is the strategically optimal choice in the majority of situations. Where this does not hold, the rational game user will spend the majority of their time on other mechanics, making the game inefficient.

4.3 Veracity

When actuating a data-providing mechanic, people choose how to actuate it: they have to pick a datum out of a set of possible datums. The rational game user will provide the datum which maximises their total utility. Where utility is constant, they will provide the datum that requires the least effort. Where effort is constant, they will provide the datum of the highest utility. This has different implications for different kinds of human subject data. Where we are interested in assessing a participant's *aptitude* as expressed in a maximum performance (such as fitness, range of vocabulary, IQ), the design is relatively straightforward: maximum performance should maximize in-game utility. The long jump in athletics is such a simple applied game for learning the maximum jumping distance of participants. The game's mechanic (jumping) is necessarily integrated with the to be provided data (jumped distance), and as its sole mechanic, it is strategically central to the game. The required datum for actuating the mechanic is a jump of any distance. If the player were only rewarded for jumping, no matter how far, they would rationally minimise their effort and jump as short as possible – they would game the system. Therefore, the veracity principle requires that the in-game utility of jumping further needs to continually increase and do so in excess of the required additional effort to motivate players to make an honest effort to jump as far as they can. The player could still jump dishonestly, but in that case would cheat themselves out of their own utility.

But what about revealing subjective attitudes, values, preferences, or spontaneous inclinations? Here the veracity requirement flips into a cautionary principle. For these kinds of data, designers need to ensure that all possible data – all possible ways of actuating the connected mechanic – are of equal overall utility, that is equally strategically worthwhile/worthless and equally effortful/effortless.

In such a case, there is no marginal utility to providing one datum over another. The reasons for choosing one over the other are therefore fully sensitive to the inclinations of the individual (or non-conscious variables under study). For instance, assume a party game designed to reveal personality traits like agreeableness. Each round, players draw a card describing a social situation and three further cards describing various personality traits (agreeable, neurotic, etc.), asking the player to choose one trait to act out. The game designer wants to assess personality by observing which of the three trait card a player spontaneously chooses to act out. Assume further that the other players decide whether to give the acting player a point for their performance or not. If they are instructed to give points based on how much they *liked* the performance, acting players will be strategically biased to choose the most likable trait to act out. If certain traits are systematically more difficult to act out than others no matter one's preference, actors will be biased not to choose those. Only if choosing one trait card over the others has no such marginal utility or disutility will it be an honest signal of the players' spontaneous inclinations.

As a rational game user, the player will only maximize marginal in-game utility to the extent that doing so doesn't incur larger marginal costs in disutility or extrinsic utility, which can be connected to honesty. E.g., if truthfully revealing one's sexual preference in a game of *Truth or Dare* is perceived to be highly socially undesirable, even if doing so would earn more points for one's team, the rational game user will be more likely to lie or choose a dare task. Contrariwise, the *alibi function* of games [24] may enable more honest responding: as players can claim plausible deniability (I didn't *want* to do it, I *had* to do it to win), this lowers the perceived disutility of acting in accord with one's spontaneous inclinations, even if they are thought to be socially undesirable.

5 EVALUATION

The most immediate use of Intrinsic Elicitation is as an heuristic evaluation tool to assess game designs for data elicitation. To demonstrate this, we will discuss *Urbanopoloy* [15], an applied game for data collection. *Urbanopoly* is a competitive multiplayer mobile game for gathering location-based data about urban landmarks, clothed in the fiction of players being rich landlords. The goal is to have the most virtual cash by buying and trading real venues on the game map that generate different daily cash bonuses every day the player logs into the game. When players open the game, they can see and click on venues in their geographic vicinity. When a venue is free, they can purchase it. Is the venue already owned, the player spins a "wheel of fortune" resulting in a venue-related task such as making an advertisement (entering venue information and shooting a photo of it), answering venue-related quiz questions, or rating photos. All these activities receive virtual cash rewards. The wheel can also trigger a chance to take the venue from the owner or paying rent to the owner. Once multiple entries are collected on a venue, a weighted majority agreement algorithm is used to identify consensus truth and trigger a cash penalty for non-consensus answers. Thus, *Urbanopoly* has nine mechanics or verbs: log in, move to venue, buy, spin wheel, answer quiz, judge photos, take venue, pay rent (figure 3).

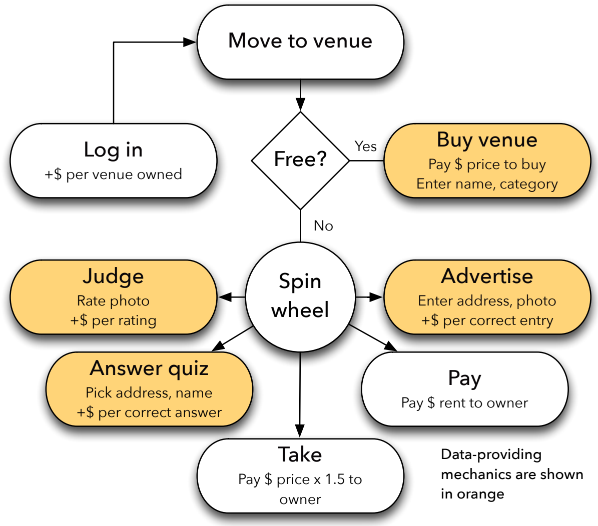

Figure 3: The core game loop of *Urbanopoly*.

5.1 Necessity

The necessity principle requires data provision to be inherent to the mechanics changing the game state. Four of *Urbanopoly*'s nine mechanics entail data provision: purchasing venues requires buyers to enter their name and category from a predefined list; advertise (enter information and photo); answer quizzes (pick the correct datum from a list of previous player answers); and rate photos on a scale of 1-5. In each case, the game state doesn't change until players enter data. Yet the data provision is sometimes weakly integrated into the mechanics: There is no diegetic or practical reason why buying a venue should entail entering its name and category – the interesting choice is which venue to purchase based on price and likely earnings. Thus, entering data during purchase does not partake in or contribute to the intrinsic utility of game enjoyment. Contrast this with answering quiz questions, where guessing the answer and providing verification data are one and the same. Rating a photo, finally, is data provision as a game mechanic, but barely an interesting core loop: there is no challenge or interesting decision to expressing one's preference, as any rating is equally rewarded. Finally, the daily log in and rent payment allow players to advance towards their game goal (earning cash) without entering any data.

5.2 Centrality

The presence of such "pure gaming features" [14] isn't problematic as long as data-providing mechanics are strategically central to winning the game, which makes it optimal to spend the majority of playtime actuating them (centrality principle). Translated into *Urbanopoly*: Are the data-providing mechanics the best time investment for virtual cash earned? Given the fact that players need to physically move to venues to actuate them, and that verifying or entering data may require checking house numbers, street names, and the like, each game action comes with non-trivial time and effort.

Here, the game shares one major downside with its inspiration, *Monopoly*: Once a player owns a range of properties, checking in daily and receiving rent payments from other players become a steady cash source compared to which the effort of data-providing actions seems hardly worth it. The less venues a player owns, the more rational it is to invest time into selecting venues and spinning the wheel, as only this allows to acquire cash through data-provision tasks (which are more likely to show on the wheel than the pay rent option) and take venues from their owner. Here, the centrality principle can guide the balancing of virtual cash rewards per mechanic: these should be arranged so that the data provision most desired by the system designer delivers the best cost/benefit ratio.

5.3 Veracity

The veracity principle states that providing the desired kind of data should have the highest utility and the least effort. Here, *Urbanopoly* faces the challenge that it cannot distinguish honest from dishonest entries at the time at which initial data is provided, as it relies on multiple entries on the same item to derive a consensus truth. This means that (in the short term), it is rational for players to game the system and enter as many data points as fast as possible regardless of their accuracy, as every entry is rewarded equally. However, players are warned that their 'karma' will penalise them if their entries are later found to be incorrect. This phrasing calls on meaningfulness and social norms, but it also refers to the fact that players whose entries lie outside the consensus will receive a cash penalty later. This is arguably a good first step to ensure veracity, but relies on players understanding said delayed penalty. However, it misses out on optimising for broad coverage as a data quality. Players receive the same amount of reward for doing the same data-providing activities repeatedly for the same places. And since venue density varies geographically, a rational player will move to and provide data on the same close-by, densely packed venues again and again.

5.4 Summary Evaluation

Analysing *Urbanopoly* through the lens of the three principles of Intrinsic Elicitation immediately foregrounded several design shortcomings. Evaluating for necessity showed that entering data is weakly integrated into the buy venue mechanic, likely causing player frustration not enjoyment. As for centrality, the more venues a player owns, the less central data-providing mechanics become to them, suggesting to rebalance the cash payouts per mechanic. The biggest takeaway came from the veracity principle. While introducing a cash penalty for non-consensus answers should prevent careless responses, the current game design makes it rational to revisit the same cluster of close-by venues constantly, rather than providing new data on venues not yet covered.

6 DISCUSSION AND CONCLUSION

We began this paper with the observation that current applied games for data collection are dominated by two basic templates – *GWAP*-style classification games and *Foldit*-style solution discovery games. For both, the literature provides a general design approach of gamification+validation: data volume is motivated with particular game design elements, data quality is then separately ensured with validation strategies. While popular, we noted that this approach falls short when it comes to human subject data, as these often involve no subject-external ground truth to validate against, which the gamification+validation approach requires. Furthermore, by making certain in-game responses more enjoyable or strategically opportune, gamification strategies may threaten validity by biasing or overshadowing 'spontaneous' expressions of preferences, attitudes, or beliefs. In response, we articulated the need for broader elicitation approaches that *integrate* motivation and data quality at once. Based on literature on data collection games, crowdsourcing, survey response, and citizen science, we developed a rational choice model explaining why players choose to provide certain data in games, the *Rational Game User Model*. Building on and incorporating Jonas Heide Smith's Rational Player Model, it predicts that players choose in-game actions providing data that maximise their overall utility, comprising three factors: extrinsic utility such as social norms and monetary incentives, extrinsic disutility such as opportunity costs and displeasure of the effort of providing the data, and intrinsic utility, comprising meaning and game enjoyment. Game enjoyment in turn is maximised by the player trying to maximise their in-game utility, i.e. choosing the course of action that maximises their odds to win. From this model, we then derived three principles for designing human subject data collection games we summarisingly call the *Intrinsic Elicitation* approach. A good data collection game integrates data generation into its enjoyable mechanics and core loops such that it is (a) necessary to actuate the mechanics, (b) strategically central to gameplay, and (c) providing spontaneous or honest data has the highest utility and lowest effort, or at least equal utility and effort compared to all other available options. Finally, we illustrated the value of our approach by using it as a heuristic evaluation tool for the existing data collection game *Urbanopoly*.

6.1 Limitations

The simplicity of our model is both strength and weakness. It allows us to generate clear predictions to test, falsify, or refine the model. For instance, it predicts that players will abandon the game when the perceived overall disutility is greater than the perceived extrinsinc and intrinsic utility. Similarly, if the perceived marginal loss of meaning due to a strategically optimal but dishonest in-game action is greater than the perceived marginal gain in game enjoyment, we predict that players would refrain from choosing it. In terms of limitations, we fully expect that in the course, several assumptions will need to be complicated. First, behavioural economics demonstrates that people's rationality is bounded by biases and heuristics [13]. Second, we know that players gain game enjoyment from more than just playing optimally: curiosity, surprise, engrossment, relatedness are other important sources [8]. For instance, a key aspect of good game design is giving players meaningful choices. Dominant strategies (where there is only one rational move) are not a concern for the rational player, but rational game users may feel like they lack autonomy or are not making any meaningful choices in this situation, reducing their overall game enjoyment. Against the centrality principle, game designers may therefore need to introduce alternative mechanics that are sometimes the preferable choice so as to make choosing the data-providing mechanic a non-trivial choice. Third, we suspect that there may be systematic causal effects between individual factors that we didn't specify – e.g., self-determination theory would predict that adding tangible rewards may undermine intrinsic game enjoyment [21]. Designers may wish to use their own more nuanced knowledge of player behaviour to violate some of the principles.

Beyond these conceptual limitations, it may not be possible for every kind of data to be integrated into a game in such a way that satisfies all aspects of intrinsic elicitation. For example, it might be technically impossible for the game to distinguish target from non-target data – in a language elicitation game, natural language processing algorithms may not be fast or sophisticated enough, for instance. Following the model, however, it is possible to suggest implications for such limit cases. In the example, we would expect noisy data as players produce both target and non-target forms. If some forms are more effortful to produce than others, we can expect that the collected data will be biased against those forms, and we may want to use statistical techniques to identify and counteract this bias.

6.2 Next Steps

The first necessary next step for the Rational Game User Model is to empirically test its predictions. Here, we consider the suggestions of Siu and colleagues instructive [73]: designing a series of studies and replications where we define and hold player experience and task completion metrics constant while varying individual factors of our model in controlled A/B design experiments, with data collection goals where the ground truth can be cross-validated against pre-existing data – for instance prior responses of participants to standard personality instruments, of replicating well-established spontaneous inclinations. A second necessary step is testing the usefulness and ease of use of the design approach of intrinsic elicitation with data collection designers. If the model and approach prove reliable and useful for human subject data collection, a third step would be to test its generalizability for data collection games like GWAPs as well.

7 ACKNOWLEDGEMENTS

This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games & Games Intelligence (IGGI) [EP/L015846/1] and the Digital Creativity Labs (digitalcreativity.ac.uk), jointly funded by EPSRC/ AHRC/Innovate UK under grant no. EP/M023265/1.

REFERENCES

[1] Luis Von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 319–326.

[2] Luis Von Ahn and Laura Dabbish. 2008. Designing games with a purpose. *Commun. ACM* 51, 8 (2008), 58–67.

[3] Luis Von Ahn, Mihir Kedia, and Manuel Blum. 2006. Verbosity: a game for collecting common-sense facts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, 75–78.

[4] Ryan S J Baker. 2011. Gaming the System: A Retrospective Look. *Philippine Computing Journal* 6, 2 (2011), 9–13.

[5] Avinoam Baruch, Andrew May, and Dapeng Yu. 2016. The motivations, enablers and barriers for voluntary participation in an online crowdsourcing platform. *Computers in Human Behavior* 64 (2016), 923–931. https://doi.org/10.1016/j.chb.2016.07.039

[6] Francesco Bellotti, Bill Kapralos, Kiju Lee, Pablo Moreno-Ger, and Riccardo Berta. 2013. Assessment in and of Serious Games: An Overview. *Advances in Human-Computer Interaction* 2013 (2013), 1–11.

[7] Gwyn Bevan and Christopher Hood. 2006. What's Measured is What Matters: Targets and Gaming the System in the English Public Health Care System. *Public Administration* 84, 3 (2006), 517–538.

[8] Elizabeth A. Boyle, Thomas M. Connolly, Thomas Hainey, and James M. Boyle. 2012. Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior* 28, 3 (may 2012), 771–780. https://doi.org/10.1016/j.chb.2011.11.020

[9] John Broome. 1991. "Utility". *Economics and Philosophy* 7 (1991), 1–12.

[10] Harriet R. Brown, Peter Zeidman, Peter Smittenaar, Rick A. Adams, Fiona McNab, Robb B. Rutledge, and Raymond J. Dolan. 2014. Crowdsourcing for cognitive science - The utility of smartphones. *PLoS ONE* 9, 7 (2014).

[11] Raymond C. Browning, Emily A. Baker, Jessica A. Herron, and Rodger Kram. 2006. Effects of obesity and sex on the energetic cost and preferred speed of walking. *Journal of Applied Physiology* 100, 2 (2006), 390–398. https://doi.org/10.1152/japplphysiol.00767.2005

[12] Bauke Buikstra, Rens Kortmann, Eric Klaassen, Laurens Rook, and Alexander Verbraeck. 2015. In Pursuit of Evidence: a Design and Empirical Study of a Gamified Online Marketplace. In *AcademicMindTrek"15*. ACM, Tampere, 73–80.

[13] Colin F. Camerer and George Loewenstein. 2004. Behavioral Economics: Past, Present, Future. In *Advances in Behavioral Economics*. Princeton University Press, Princeton, Oxford, 2–51.

[14] I Celino and D Cerizza. 2012. Urbanopoly: collection and quality assesment of geo-spatial linked data via a human computation game. In *Proceedings of the 10th Semantic Web Challenge*. 148–163.

[15] I Celino, D Cerizza, S Contessa, M. Corubolo, D. Dell'Aglio, E. D. Valle, and S Fumeo. 2012. Urbanopoly–A Social and Location-Based Game with a Purpose to Crowdsource Your Urban Data. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. IEEE, 910–913.

[16] Otto Chrons and Sami Sundell. 2011. Digitalkoot: Making Old Archives Accessible Using Crowdsourcing. In *Human Computation: Papers from the 2011 AAAI Workshop*. 20–25.

[17] Seth Cooper. 2015. Massively Multiplayer Research: Gamification and (Citizen) Science. In *The Gameful World: Approaches, Issues, Applications*, Steffen P. Walz and Sebastian Deterding (Eds.). MIT Press, Cambridge, MA, 487–500.

[18] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, and Foldit Players. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466, 7307 (aug 2010), 756–60.

[19] Seth Cooper, Adrien Treuille, Janos Barbero, Andrew Leaver-Fay, Kathleen Tuite, Firas Khatib, Alex Cho Snyder, Michael Beenen, David Salesin, David Baker, and Zoran Popović. 2010. The Challenge of Designing Scientific Discovery Games. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*. 40–47.

[20] Vickie Curtis. 2015. Motivation to Participate in an Online Citizen Science Game: A Study of Foldit. *Science Communication* 37, 6 (2015), 723–746.

[21] Edward L Deci and Richard M Ryan. 2000. The âĂIJWhatâĂİ and âĂIJWhyâĂİ of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry* 11, 4 (2000), 227–268.

[22] Sebastian Deterding. 2014. *Modes of Play: A Frame Analytic Account of Video Game Play*. Doctoral. Hamburg University.

[23] Sebastian Deterding. 2015. The Lens of Intrinsic Skill Atoms: A Method for Gameful Design. *Human-Computer Interaction* 30, 3-4 (2015), 294–335.

[24] Sebastian Deterding. 2018. Alibis for Adult Play: A Goffmanian Account of Escaping Embarrassment in Adult Play. *Games and Culture* 13, 3 (2018), 260–279. https://doi.org/10.1177/1555412017721086

[25] S. Deterding, A. Canossa, C. Harteveld, S. Cooper, L.E. Nacke, and J.R. Whitson. 2015. Gamifying research: Strategies, opportunities, challenges, ethics. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 18.

[26] Sebastian Deterding, Anders Drachen, Catherine Flick, Ben Medler, Lies van Roessel, and José P. Zagal. 2018. The Coming Sh*t Storm? Game Research and Design Ethics after Facebook. In *FDG'18*. Malmö.

[27] Sam Devlin, Peter I Cowling, Daniel Kudenko, Nikolaos Goumagias, Alberto Nucciareli, Ignazio Cabras, Kiran Jude Fernandes, and Feng Li. 2014. Game Intelligence. In *Computational Intelligence and Games*. IEEE, Dortmund, 1–8.

[28] Alejandro Echeverría, Enrique Barrios, Miguel Nussbaum, Matías Améstica, and Sandra Leclerc. 2012. The Atomic Intrinsic Integration Approach: A structured methodology for the design of games for the conceptual understanding of physics. *Computers & Education* 59, 2 (2012), 806–816.

[29] Weimiao Fan and Zheng Yan. 2010. Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior* 26, 2 (2010), 132–139.

[30] Mary Flanagan and Sukdith Punjasthitkul. 2013. Citizen Archivists at Play: Game Design for Gathering Metadata for Cultural Heritage Institutions. In *DiGRA Conference*. Proceedings of DiGRA 2013.

[31] Cyrus K. Foroughi, Carolyn Serraino, Raja Parasuraman, and Deborah A. Boehm-Davis. 2016. Can we create a measure of fluid intelligence using Puzzle Creator within Portal 2? *Intelligence* 56 (2016), 58–64.
[32] Luca Galli. 2014. Matching Game Mechanics and Human Computation Tasks in Games with a Purpose. In *Proceedings of the 2014 ACM International Workshop on Serious Games*. 9–14.
[33] K. Gary, R. Stoll, P. Rallabhandi, M. Patwardhan, D. Hamel, A. Amresh, A. Pina, K. Cleary, and Z. Quezado. 2017. MHealth games as rewards: Incentive or distraction? *ACM International Conference Proceeding Series* Part F1286 (2017), 7–10.
[34] Jacqueline Gaston and Seth Cooper. 2017. To Three or not to Three: Improving Human Computation Game Onboarding with a Three-Star System. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* (2017), 5034–5039.
[35] H. Geoghegan, A. Dyke, R. Pateman, S. West, and G. Everett. 2016. *Understanding Motivations for Citizen Science. Final Report on behalf of the UK Environmental Observation Framework (UKEOF).* Technical Report. 124 pages. http://www.ukeof.org.uk/resources/citizen-science-resources/MotivationsforCSREPORTFINALMay2016.pdf
[36] Erving Goffman. 1972. *Encounters: Two Studies in the Sociology of Interaction*. Penguin, Harmondsworth.
[37] Dion H. Goh and Chei Sian Lee. 2011. Perceptions, quality and motivational needs in image tagging human computation games. *Journal of Information Science* 37, 5 (2011), 515–531.
[38] Dion Hoe Lian Goh and Chei Sian Lee. 2011. Understanding playability and motivational needs in human computation games. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 7008 LNCS. 108–117.
[39] Dion Hoe Lian Goh, Ei Pa Pa Pe-Than, and Chei Sian Lee. 2017. Perceptions of virtual reward systems in crowdsourcing games. *Computers in Human Behavior* 70 (2017), 365–374.
[40] Peter M Gollwitzer and Gabriele Oettingen. 2012. Goal Pursuit. In *The Oxford Handbook of Human Motivation*, Richard M. Ryan (Ed.). Oxford University Press, Oxford, 208–231.
[41] M P Jacob Habgood and Shaaron E Ainsworth. 2011. Motivating Children to Learn Effectively : Exploring the Value of Intrinsic Integration in Educational Games. *The Journal of the Learning Sciences* 20, August (2011), 169–206.
[42] Kieran Healy. 2017. Fuck Nuance. *Sociological Theory* 35, 2 (2017), 118–127.
[43] Richard Heeks. 2010. Understanding "Gold Farming" and Real-Money Trading as the Intersection of Real and Virtual Economies. *Journal of Virtual Worlds Research* 2, 4 (2010), 1–27.
[44] Peter Jamieson, Jack Hall, and Lindsay Grace. 2012. Research Directions for Pushing Harnessing Human Computation to Mainstream Video Games. In *Meaningful Play 2012*. East Lansing, MI.
[45] John A. Johnson. 2005. Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality* 39, 1 SPEC. ISS. (2005), 103–129.
[46] Jesper Juul. 2005. *Half-Real: Video Games between Real Rules and Fictional Worlds*. MIT Press, Cambridge, MA.
[47] Jesper Juul. 2008. The Magic Circle and the Puzzle Piece. In *Conference Proceedings of the Philosophy of Computer Games 2008*. 56–67.
[48] Florian Keusch and Chan Zhang. 2017. A Review of Issues in Gamified Surveys. *Social Science Computer Review* 35, 2 (2017), 147–166.
[49] Dominik J. Leiner. 2013. Too Fast, Too Straight, Too Weird: Post Hoc Identification of Meaningless Data in Internet Surveys. *SSRN Electronic Journal* November 2013 (2013). https://doi.org/10.2139/ssrn.2361661
[50] Robert V Levine and Ara Norenzayan. 1999. The pace of fife in 31 countries. *Journal of Cross-Cultural Psychology* 30, 2 (1999), 178–205. https://doi.org/10.1177/0022022199030002003
[51] TW Malone. 1980. What makes things fun to learn? Heuristics for designing instructional computer games. In *Proceedings of the 3rd ACM SIGSMALL symposium*. ACM, New York, 162–169.
[52] Donald L Mccabe, Linda Klebe Treviño, and Kenneth D Butterfield. 2001. Cheating in Academic Institutions: A Decade of Research. *Ethics & Behavior* 11, 3 (2001), 219–232.
[53] Samuel Messick. 1995. Validity of Psychological Assessment: Validation of Inferences From Person's Responses and Performances as Scientific Inquiry Into Score Meaning. *American Psychologist* 50, 9 (1995), 471–749.
[54] Benedikt Morschheuser, Juho Hamari, Jonna Koivisto, and Alexander Maedche. 2017. Gamified crowdsourcing: Conceptualization, literature review, and future agenda. *International Journal of Human Computer Studies* 106, May (2017), 26–43.
[55] Martin J. Osborne and Ariel Rubinstein. 1994. *A Course in Game Theory*. MIT Press, Cambridge, MA, London.
[56] Ei Pa Pa Pe-Than, Dion Hoe Lian Goh, and Chei Sian Lee. 2012. A survey and typology of human computation games. *Proceedings of the 9th International Conference on Information Technology, ITNG 2012* (2012), 720–725.
[57] Ei Pa Pa Pe-Than, Dion Hoe Lian Goh, and Chei Sian Lee. 2012. A survey and typology of human computation games. *Proceedings of the 9th International Conference on Information Technology, ITNG 2012* (2012), 720–725. https://doi.org/10.1109/ITNG.2012.124
[58] Nathan Prestopnik, Kevin Crowston, and Jun Wang. 2014. Exploring Data Quality in Games With a Purpose. *iConference 2014 Proceedings* 2008 (2014), 1–15. https://doi.org/10.9776/14066
[59] Nathan R. Prestopnik and Jian Tang. 2015. Points, stories, worlds, and diegesis: Comparing player experiences in two citizen science games. *Computers in Human Behavior* 52 (2015), 492–506. https://doi.org/10.1016/j.chb.2015.05.051
[60] Alexander J Quinn and Benjamin B Bederson. 2011. Human Computation: A Survey and Taxonomy of a Growing Field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1403–1412.
[61] Stuart Reeves and Scott Sherwood. 2010. Five design challenges for human computation. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries - NordiCHI '10*.
[62] Jakob Rogstadius, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. 2011. An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media An*. AAAI Press, 321–328.
[63] Joel Ross and Bill Tomlinson. 2010. How games can redirect humanity's cognitive surplus for social good. *Computers in Entertainment* 8, 4 (dec 2010), Art. 25.
[64] Anurag Sarkar, Michael Williams, Sebastian Deterding, and Seth Cooper. 2017. Engagement Effects of Player Rating System-Based Matchmaking for Level Ordering in Human Computation Games. In *Proceedings of FDG'17*. ACM Press, New York.
[65] Jesse Schell. 2009. *The Art of Game Design: A Book of Lenses*. Morgan Kaufman, Amsterdam et al.
[66] Ralf Schmidt, Katharina Emmerich, and Burkhard Schmidt. 2015. Applied Games - In Search of a New Definition. In *Entertainment Computing - ICEC 2015*, K Chorianopoulos (Ed.), Vol. 9353. 100–111.
[67] Karen Schrier. 2017. What's in a name? Naming Games that Solve Real-World Problems. *Proceedings of the International Conference on the Foundations of Digital Games - FDG '17* (2017).
[68] John Scott. 2000. Rational Choice Theory. In *Understanding Contemporary Society: Theories of The Present*, G. Browning, A. Halcli, and F. Webster (Eds.). Sage, 671–85.
[69] Miguel Sicart. 2008. Defining Game Mechanics. *Game Studies* 8, 2 (2008), n.p.
[70] Miguel Sicart. 2014. *Play Matters*. MIT Press, Cambridge, MA.
[71] Miguel Sicart. 2015. Loops and Metagames: Understanding Game Design Structures. In *Foundations of Digital Games*.
[72] Kristin Siu and Mark O Riedl. 2016. Reward Systems in Human Computation Games. *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (2016), 266–275.
[73] Kristin Siu, Alexander Zook, and Mark O. Riedl. 2017. A Framework for Exploring and Evaluating Mechanics in Human Computation Games. In *FDG'17*. ACM Press, New York.
[74] Karin Slegers, Bernhard Maurer, Lizzy Bleumers, Alina Krischkowsky, Pieter Duysburgh, and Mark Blythe. 2016. Game-based HCI Methods: Workshop on Playfully Engaging Users in Design. In *CHI Extended Abstracts on Human Factors in Computing Systems*. ACM Press, New York, 3484–3491.
[75] Jonas Heide Smith. 2006. *Plans and Purposes*. Ph.D. Dissertation. The IT University of Copenhagen.
[76] David A. Spencer. 2003. Love's labor's lost? The disutility of work and work avoidance in the economic analysis of labor supply. *Review of Social Economy* 61, 2 (2003), 235–250+273.
[77] Kurt Squire. 2006. From content to context: Videogames as designed experience. *Educational researcher* 35, 8 (2006), 19–29. http://edr.sagepub.com/content/35/8/19.short
[78] Ramine Tinati, Markus Luczak-Roesch, Elena Simperl, and Wendy Hall. 2017. An investigation of player motivations in Eyewire, a gamified citizen science project. *Computers in Human Behavior* 73 (2017), 527–540.
[79] Kathleen Tuite. 2014. GWAPs: Games with a Problem. In *Foundations of Digital Games 2014*.
[80] Vanessa Torres van Grinsven. 2015. *Motivation in Business Survey Response Behaviour. Influencing motivation to improve survey outcome.*
[81] Xiaohui Wang, D.H.-L. Goh, E.-P. Lim, and E.-P. Vu. 2014. Player acceptance of human computation games: An aesthetic perspective. In *16th International Conference on Asia-Pacific Digital Libraries, ICADL 2014*, Vol. 8839. 233–242.
[82] Marika Wenemark, Andreas Persson, Helle Noorlind Brage, Tommy Svensson, and Margareta Kristenson. 2011. Applying Motivation Theory to Achieve Increased Respondent Satisfaction, Response Rate and Data Quality in a Self-administered Survey. *Journal of Official Statistics* 27, 2 (2011), 393–414.
[83] Kevin Werbach and Dan Hunter. 2012. *For the Win: How Game Thinking Can Revolutionize Your Business*. Wharton Digital Press, Philadelphia, PA.