NBER WORKING PAPER SERIES

STORM CROWDS: EVIDENCE FROM ZOONIVERSE ON CROWD CONTRIBUTION
DESIGN

Sandra Barbosu
Joshua Gans

Storm Crowds: Evidence from Zooniverse on Crowd Contribution Design
Sandra Barbosu and Joshua Gans
NBER Working Paper No. 23955
October 2017
JEL No. H42,O31

ABSTRACT

Crowdsourcing - a collaborative form of content production based on the contributions of large groups of individuals - has proliferated in the past decade. Due to this growth, recent research has focused on understanding the factors that affect its sustainability. Prior studies have highlighted the importance of volunteers' prosocial motivations, the sense of belonging to a community, and symbolic rewards within crowdsourcing websites. One factor that has received limited attention in the existing literature is how the design of crowdsourcing platforms affects their sustainability. We study whether the design element - particularly, the divisibility of contributions (i.e. whether contributing tasks are bundled together or can be carried out separately) - is a factor that affects the level and quality of crowdsourcing contributions. We investigate this in the context of Zooniverse, the world's largest crowd-sourced science site, in which volunteers contribute to scientific research by performing data processing tasks. Our choice of empirical setting is motivated by the fact that one of the Zooniverse projects, Cyclone Center, underwent a format change that decreased the divisibility of contributions, by bundling together two tasks that were previously separate. We refer to contributions for which both tasks were done as complete, and contributions for which only one task was done as incomplete. In this context, we develop a theoretical model that predicts (i) a positive relationship between contribution divisibility and the total number of contributions (i.e. complete and incomplete) per volunteer, (ii) an ambiguous relationship between contribution divisibility and the number of complete contributions per volunteer, and (iii) an ambiguous relationship between contribution divisibility and the value of complete contributions. We test these predictions empirically by exploiting the format change in Cyclone Center. We find that after the format change, which decreased contribution divisibility, (i) the total number of contributions per volunteer decreased, (ii) the number of complete contributions made by anonymous volunteers increased, while that made by registered volunteers remained unchanged, and (iii) the value of complete contributions increased because anonymous volunteers, who increased their number of complete contributions, contributed high quality contributions. Our results have strategic implications for crowdsourcing platforms because they suggest that the design of crowdsourcing platforms, specifically the divisibility of contributions, is a factor that matters for their sustainability.

Sandra Barbosu
Rotman School of Management, University of Toronto
105 St. George St.
Toronto, ON, M5S3E6
Canada
sandra.barbosu10@rotman.utoronto.ca

Joshua Gans
Rotman School of Management
University of Toronto
105 St. George Street
Toronto ON M5S 3E6
CANADA
and NBER
joshua.gans@gmail.com

1 Introduction

Crowdsourcing - a collaborative form of content production based on the contributions of crowds of people - has proliferated in the past decade, enabled by mass Internet adoption and digitization. The term *crowdsourcing* was coined by Jeff Howe, editor of *WIRED* magazine, who described it as "the act of taking a task once performed by an employee and outsourcing it to a large, dispersed, undefined group of non-experts." (Howe 2006)

Crowdsourcing is employed in a wide range of areas, including business, government/non-profits and science. Some companies use crowdsourcing for idea generation (Bayus 2013, Huang et al. 2014). Examples include *Dell's IdeaStorm Community*, which solicits new product ideas from customers (Bayus 2013), *My Starbucks Idea*, which uses customer comments to develop new coffee flavors, and *Lego Ideas*, which asks customers for new LEGO set design ideas. Other companies employ crowds to help them find unique solutions to highly challenging problems (Boudreau and Lakhani 2013). In government, initiatives like *Better Reykjavik*, which asks people to debate and prioritize the issues that will improve their city, *The Cairo Transport App Challenge*, which seeks solutions from Cairo residents to city traffic congestion, and *Future Melbourne*, which consults its citizens on city planning, show how governments can make use of crowdsourcing to actively engage citizens. In the area of non-profits, a classic example of crowdsourcing is *Wikipedia*, the free online collaborative encyclopedia, owned by the non-profit organization *Wikimedia*.

In recent years, crowdsourcing initiatives have also been used to contribute to scientific research or solve complex science problems (Lakhani et al. 2007). The world's largest crowdsourced science initiative is Zooniverse, which was founded in 2009. As of 2014, the platform had more than one million volunteers working on 59 science projects across fields.[1] Zooniverse volunteers contribute to scientific research by performing data processing and

---

[1] Details at: https://blog.zooniverse.org/2014/02/14/one-million-volunteers/

analysis tasks. In addition to Zooniverse, other crowdsourced science platforms include *The Polymath Project*, in which volunteers solve mathematics problems, *Eyewire*, where individuals do 3D puzzles to help neuroscientists understand how the brain processes visual information, and *CosmoQuest*, where people assist in identifying planet features from NASA images.

Given the recent proliferation of crowdsourcing, scholarly attention has turned to evaluating the quality of crowdsourced content, and to understanding the factors that contribute to the sustainability and efficacy of crowdsourcing initiatives. One factor that has received limited attention in the existing literature is how the design of crowdsourcing platforms affects their sustainability. For instance, one neglected feature of Wikipedia is the fact that what constitutes a contribution is undefined. Contributors can do anything from fixing typos to providing complete articles. Indeed, the term 'wiki' refers precisely to this feature of the site. By contrast, entries to Encyclopedia Britannica were typically sourced with little to no monetary reward from experts but those experts were required to provide a complete article (Greenstein and Zhu 2017). This has motivated our study into whether this design component – notably, the divisibility of contributions (i.e., whether contributing tasks are bundled together or can be carried out selectively) – is a factor in the level and quality of contributions in crowdsourcing.

We investigate this question in the context of Zooniverse, which offers us the opportunity to observe the level and quality of volunteer contributions under platform designs with varying contribution divisibility. This is an appropriate setting because one Zooniverse project, called Cyclone Center, underwent a format change that *decreased* contribution divisibility, by bundling together two tasks that could previously be done selectively. We refer to contributions in which both tasks are done as *complete*, and contributions in which only one task is done as *incomplete*. In this context, we develop a theoretical model that generates three main predictions: (i) a positive relationship between contribution divisibility and the total

number of contributions (i.e. complete and incomplete) per volunteer, (ii) an ambiguous relationship between contribution divisibility and the number of complete contributions per volunteer, and (iii) an ambiguous relationship between contribution divisibility and the value of complete contributions.

We test these predictions through an empirical approach that exploits the format change in Cyclone Center. We find that after the format change, which decreased contribution divisibility, (i) the total number of contributions per volunteer decreased, (ii) the number of complete contributions made by anonymous volunteers increased, while that made by registered volunteers remained unchanged, and (iii) the value of complete contributions increased because anonymous volunteers, who increased their number of complete contributions, contributed high quality contributions. Our findings have strategic implications for crowdsourcing platforms because they suggest that their contribution design, and specifically the divisibility of contributions, matters for the quality and level of their volunteers' participation.

1.1 Related Research

Our paper seeks to contribute to research focused on evaluating the quality of crowd-sourced content and on understanding the factors that contribute to the sustainability of crowdsourcing platforms (see, e.g. Zhao and Zhu 2014 for a review). An important factor that distinguishes crowdsourcing from production inside a firm is the absence of financial incentives. Therefore, understanding volunteers' motivations to contribute, and other factors affecting contribution levels, is critical for the sustainability of crowdsourcing initiatives (Butler 2001).

Several studies have found that prosocial motives (i.e. a desire to benefit others) are prevalent among crowdsourcing volunteers. Schroer and Hertel (2009) found that highly engaged Wikipedians are intrinsically motivated, rather than by self-interest. Rashid et

al. (2006) showed that informing volunteers of the value of their contribution increases contribution levels. Peddibhotla and Subramani (2007) further found that prosocial motives positively affect contribution quality.

One factor that has received limited attention in the existing research is how the design of crowdsourcing platforms affects volunteer contributions; in this paper, we bring the focus to this question. The evidence of prosocial motives among crowdsourcing volunteers found by prior studies would suggest that volunteers care about the overall contribution output, implying that the contribution design would matter less, as long as it leads to higher output. However, our empirical findings in Section 4 show that when contribution divisibility decreases, volunteers respond by decreasing the total number of contributions they make. This finding is not surprising given the public good nature of crowdsourcing, but it is in contradiction to hypotheses that volunteers have purely prosocial motives. Instead, even though rewards are non-monetary, basic economic principles still inform how to generate more contributions. Indeed, Gallus (2016) finds that symbolic rewards are important for the retention of new Wikipedia volunteers, suggesting that awards can be effective even if they have no external benefits outside the crowdsourcing community. This may be due to the social effects of such communities, documented by Zhang and Zhu (2011).

Our study, investigating the importance of contribution design for the level and quality of volunteer contributions, is most closely related to research by Aaltonen and Seiler (2016). They explore how the cumulative nature of content production on Wikipedia affects contribution levels, and find that longer articles lead to significantly higher contribution levels. The authors propose several explanations that could account for this finding: (i) building on existing content rather than creating a whole article lowers the cost of editing, thereby making small edits valuable; (ii) existing content provides new information about a topic; and/or (iii) existing content makes incomplete parts of the article more salient to the volunteers. However, they do not explore which mechanism is responsible. In our paper, we

investigate the effect of the first mechanism: how the divisibility of contributions (i.e. the cost of contributing) affects volunteer contributions.

The paper proceeds as follows. Section 2 presents our theoretical model and its empirical predictions. Section 3 describes the empirical setting and data. Section 4 presents our empirical approach and results. Section 5 concludes.

2   Formal Model

Here we provide a model based on Zooniverse crowdsourced science to understand the impact of contribution divisibility in the context of public good provision. Standard models of public goods tend to trade-off free riding (that is, you want others to incur the costs of providing public goods) versus intrinsic motivation (that is, some individuals may intrinsically value providing public goods). In those models, the main outcome of interest is the level of public good provision. We share that interest in the model provided here. However, what we are looking to examine is how design elements in the contribution process – namely, whether contributing tasks are bundled together or can be selectively carried out – impacts the overall level of public good provision. Divisibility has not been, to our knowledge, theoretically examined in the literature, so building a model to demonstrate its impact is instructive.

2.1   Model Set-up

Suppose there are n symmetric agents (crowdsourcing volunteers). Each agent has the ability to provide a public good, which we think of as a completed crowd science contribution. We assume that the contribution is comprised of two tasks. The first task involves a (private) cost of c and, if it is the only task done, results in value of v. The second task can only be done if the first task is complete. It can improve value by *(V-v)* - making the value of a complete contribution, V, and an incomplete contribution, v. There is uncertainty regarding the cost

associated with the second task. First, there is common uncertainty (across individuals) where the costs could be high, i.e. c_H, with probability γ or low, i.e. $c_L < c_H$, with probability 1-γ. Second, there is individual uncertainty (specific to individuals) where the costs might be lower for some individuals at a point in time. With probability α, the costs associated with the second task are a fraction, $a \in [0,1]$ of c_L and c_H, as the case may be. Otherwise, with probability $1 - \alpha$, the costs are simply c_L and c_H.

A designer can choose a technology that dictates how contributions can be made. One option is for the design to be *non-divisible*. In that case, any agent who elects to make a contribution is the only agent who can complete the contribution. An alternative option is for the design to be *divisible*. In that case, even if one agent conducts partial contributions, their contribution is accepted and has some value. We will compare the designer's choices among these alternatives.

Non-Divisible Case

To start, suppose that any agent who chooses to contribute must do both tasks. The timing of the (indivisible) crowd contribution game is as follows:

1. Nature selects an agent at random.

2. Individual uncertainty is resolved for the selected agent (that is, the agent learns if second task costs are discounted by a or not).

3. The selected agent then chooses whether to undertake both tasks or not.

   - If the agent does the tasks, the period ends and payoffs are realized with the contributing agent receiving V less their costs and all other agents receiving V.

   - If the agent only completes the first task, no value is realized and no contribution is registered.

- If the agent chooses not to contribute, no agent receives a payoff.

4. The period ends with a discount factor (common to all agents) of δ applied.

In this structure, a critical assumption is that agents do not know their own individual cost uncertainty realization until they are selected. This corresponds to the notion that individual costs differ from period to period. If this assumption did not hold, this would complicate the analysis as we would have to track two agent types over time. However, we do not believe this provides any insight to justify that complexity. The other important assumption is that the agent does not know the common cost uncertainty resolution when committing to undertake the contribution. This assumption is reasonable given the empirical application to Zooniverse that does not disclose the extent of analysis required when an agent chooses to undertake a task.

We make the following assumptions to focus on cases of interest. First, contributing is efficient even for agents with high cost outcomes. That is,

$$V > c + c_H$$

This implies that an agent who has completed the first task will, in the indivisible design case, choose to complete the second task.

Let $\pi(1)$ be the discounted payoff to an agent who is selected to contribute and $\pi(0)$ be the discounted payoff to any agent not selected. Also let q denote the probability that a selected agent chooses to contribute. Then,

$$\pi(1) = q(V - C) + (1 - q)\delta(\frac{1}{n}\pi(1) + \frac{n-1}{n}\pi(0)),$$

$$\pi(0) = qV + (1 - q)\delta(\frac{1}{n}\pi(1) + \frac{n-1}{n}\pi(0))$$

$$\text{where } C \equiv c + \alpha a(\gamma c_H + (1-\gamma)c_L) + (1-\alpha)(\gamma c_H + (1-\gamma)c_L)$$

We look for symmetric, static Nash equilibria in this game. The equilibrium q is found at the point of indifference a selected agent would have between contributing and not contributing. That is, it is the q that solves:

$$V - C = \delta(\frac{1}{n}\pi(1) + \frac{n-1}{n}\pi(0))$$

Solving this equation for q yields:

$$\hat{q} = min\{1, \frac{n(1-\delta)(V-C)}{(n-1)\delta C}\}$$

Thus, if $n = 1$, it is easy to see that contributions will always occur (i.e., $\hat{q} = 1$) and that $\hat{q}$ is decreasing in n. This reflects the standard intuition for free-riding in public goods games.

Divisible Case

We now amend the crowd contribution game to allow for the possibility that the agent can generate value by doing just the first task even if the second is incomplete.

1. Nature selects an agent at random.

2. Individual uncertainty is resolved for the selected agent (that is, the agent learns if second task costs are discounted by a or not).

3. The selected agent then chooses whether to undertake the first task.

   - If the agent does the task, a payoff of v is realized for all agents and the agent can choose to do the second task (i.e., we move to 4)

   - If the agent chooses not to contribute, no agent receives a payoff.

9

4. If the agent has done the first task, second task common cost uncertainty is resolved and the agent chooses whether to do the second task.

   - If the agent does the task, a payoff of V is realized.

   - If the agent does not do the second task, no agent receives a further payoff that period.

5. The period ends with a discount factor (common to all agents) of δ applied.

In this set-up, it is possible for the agent to contribute by doing one task – although the value realized is less than that of a complete contribution. At step 4, the agent will always choose to do the second task if:

$$V - v < c_H$$

In this case, the outcome will be the same as that for the indivisible case. Therefore, to make things interesting we assume the following:

$$c_H > V - v > max[ac_H, c_L] + \tilde{V} - C$$

Under this assumption, at step 4, agents with a high (common and individual) cost realization will choose to stop there and leave the contribution incomplete. Otherwise, they will complete the contribution as delay would be too costly even when there is some intermediate value created from an incomplete contribution.

Suppose that a selected agent chooses to contribute with probability $\tilde{q}$. Then,

$$\pi(1) = \tilde{q}(\tilde{V} - \tilde{C}) + (1 - \tilde{q})\delta(\frac{1}{n}\pi(1) + \frac{n-1}{n}\pi(0)),$$

$$\pi(0) = \tilde{q}\tilde{V} + (1 - q)\delta(\frac{1}{n}\pi(1) + \frac{n-1}{n}\pi(0))$$

where $\tilde{C} \equiv c + \alpha a(\gamma c_H + (1-\gamma)c_L) + (1-\alpha)(1-\gamma)c_L$ and $\tilde{V} \equiv V - (1-\alpha)\gamma(V-v)$

We look for symmetric, static Nash equilibria in this game. The equilibrium $\tilde{q}$ is found at the point of indifference a selected agent would have between contributing and not contributing. That is, it is the $\tilde{q}$ that solves:

$$\tilde{V} - \tilde{C} = \delta(\frac{1}{n}\pi(1) + \frac{n-1}{n}\pi(0))$$

Solving this equation for $\tilde{q}$ yields:

$$\tilde{q} = min\{1, \frac{n(1-\delta)(\tilde{V}-\tilde{C})}{(n-1)\delta\tilde{C}}\}$$

Once again, if $n = 1$, it is easy to see that contributions will always occur (i.e., $\tilde{q} = 1$) and that $\tilde{q}$ is decreasing in n.[2]

Comparison

We are now in a position to compare the indivisible and divisible cases. First, looking at the rate of contribution, note that:

$$\tilde{q} > \hat{q} \Rightarrow \frac{n(1-\delta)(\tilde{V}-\tilde{C})}{(n-1)\delta\tilde{C}} > \frac{n(1-\delta)(V-C)}{(n-1)\delta C} \Rightarrow \frac{\tilde{V}-\tilde{C}}{\tilde{C}} > \frac{V-C}{C}$$

which holds as $c_H > V - v$. In effect, the payoff to contributions by individuals is higher in the divisible case because contributions are not 'forced' when costs are too high.

Next, we can look at what happens to the number of complete contributions.

$$\tilde{q}(1-(1-\alpha)\gamma) > \hat{q} \implies \frac{\tilde{V}-\tilde{C}}{\tilde{C}}(1-(1-\alpha)\gamma) > \frac{V-C}{C}$$

This is ambiguous. Put simply, while being forced to provide only complete contributions causes agents expecting lower costs to do so, that compulsion may cause a reduction in those

---
[2] A proof of this claim is provided in Appendix A.

contributing at all.

Finally, if the design goal is to maximize complete contribution value, then we are interested in comparing $\tilde{q}\tilde{V}$ and $\hat{q}V$.

$$\tilde{q}\tilde{V} > \hat{q}V \implies \frac{\tilde{V} - \tilde{C}}{\tilde{C}}\tilde{V} > \frac{V - C}{C}V$$

Again this is ambiguous, as $\tilde{V} < V$. However, if c_H is high and a is low, then we can show that the value of completed analysis falls as we move to divisible contributions.

2.2 Implications for Empirical Analysis

Our model generates three main empirical predictions that we test using data from Zooniverse. The first is that an editor's probability of contributing total edits will be higher in a divisible than in a non-divisible editing format.

The second is that, while it is ambiguous whether an editor's probability of contributing complete edits will be higher or lower in a divisible or a non-divisible format, with further assumptions a clearer prediction can be made. For instance, suppose we can identify a class of volunteers who have more experience with the platform and know their own contribution costs. Or to put this another way, suppose that there was a class of volunteers with a lower a. In this situation, the model implies that this class of volunteers is likely to contribute complete contributions even when the format is indivisible. Consequently, compared with other classes, we would expect that their rate of complete contributions would be unchanged as the platform moves between indivisible and divisible contribution formats. Below, we argue that registered volunteers (i.e. those who make contributions while logged in to a Zooniverse account), may constitute such a 'low cost' class, and examine how their contributions differ from those of anonymous volunteers.

Finally, an assumption in the model is that contributors care about overall contributions

and not their specific contributions. This is what gives rise to the public good effects in the model. These characteristics of agent utility can be tested by examining whether the size of the crowd matters for contributions. From a modeling perspective, the same comparative statics on contributions hold if contributing has a purely private benefit. We have presented it here as a public good contribution model to provide a more general context. However, the main design predictions are robust to alternative specifications.

3 Empirical Setting and Data

The empirical setting in which we test our model's theoretical predictions is the crowd-sourced science platform Zooniverse. In particular, we focus on one of its projects, Cyclone Center, which allows us to observe volunteer contributions under designs with varying contribution divisibility.

3.1 Empirical Setting: Zooniverse

Originally launched in December 2009 with one project, Zooniverse has grown to be the world's largest crowdsourced science platform, containing 59 science projects in eleven disciplines, ranging from natural sciences, to social sciences and humanities. To date, the efforts of volunteers have resulted in numerous scientific discoveries and more than 100 academic publications. The success of Zooniverse relies on the sustained contributions of volunteers, who contribute to scientific research by performing data analysis and processing tasks. Depending on the project, the tasks may consist of classifying, marking, identifying, transcribing, or pattern matching the contents of an image data file. For each project, volunteers are guided through a brief tutorial that teaches them how to properly contribute (Simpson 2014). Volunteers can contribute either anonymously, or by registering for a Zooniverse account, and contributing while logged in. Zooniverse volunteers are known as 'editors', so we refer to

them in this way in the paper.

3.1.1 Cyclone Center, a Zooniverse Project

The goal of Cyclone Center is to assist climatologists studying tropical storms, in order to more accurately predict the behavior of future storms. Editors are shown infrared satellite images of storms, and are asked to analyze these images by performing several tasks, such as pattern matching the type of storm (i.e. determine its intensity), and identifying various attributes. In Cyclone Center, as in all other Zooniverse projects, the same image is shown to multiple editors (typically ten) in order to improve the accuracy of the answers. However, an editor cannot see how others have previously analyzed the storm image. Two features of Cyclone Center make this project an appropriate empirical setting in which to test our model's theoretical predictions.

Feature 1: Cyclone Center format change (decrease in contribution divisibility)

In June 2013, nine months after the launch of Cyclone Center, the project underwent a format change that decreased the divisibility of contributions. Before the format change, each storm image had three questions that editors were required to answer in order to submit an edit, and between two and five questions that were optional. The optional questions asked for more detailed information about the storm. If editors chose not to answer the optional questions, they could still submit an edit as long as they answered the three required questions. The three required questions asked editors to identify (i) the storm type (a reflection of its intensity), (ii) the storm subtype, and (iii) whether the storm's intensity was increasing or decreasing compared to an earlier image of the storm. The number of optional questions varied depending on the type of storm identified by the editor in the second required question.

A storm's type captures its intensity, and varies over the lifetime of the storm. At any

given time, storm intensity can be one of four types: shear, embedded, curved or eye. This classification is based on the Dvorak Technique, a method used to estimate storm intensity based on cloud patterns, which was originally developed in the 1970s and is still widely used today (Dvorak 1973). Figure 1 shows the Dvorak classification technique: the four possible storm types and their associated cloud patterns.

———————————————————

Insert Figure 1 about here

———————————————————

On June 13, 2013, the format change was implemented, and was announced on the project's blog. Based on discussions with Cyclone Center staff, plans for the change were not announced prior to the change taking effect, so it was unanticipated by editors. We refer to the pre-format change period as Version 1, and the post-format change period as Version 2 of the project. Essentially, the format change decreased the divisibility of contributions by bundling together the two tasks (i.e. required and optional questions) from Version 1.

The format change allows us to observe editor contributions in two editing designs with varying contribution divisibility. We employ a difference-in-differences (DD) approach that exploits the format change, and compare contribution levels in Cyclone Center to those in a similar project, Galaxy Zoo, which did not undergo any format change. We describe the DD strategy and its key identifying assumption in Section 4.1.

Feature 2: Variation in contribution divisibility across storm types

The second feature that makes Cyclone Center a good empirical setting for testing our theoretical predictions is the project's format during Version 1 (i.e. the first nine months prior to the format change). In Version 1, each storm image had three required questions, and between two and five optional questions, depending on the type of storm identified by

the editor, as follows: shear and embedded storm types had two optional questions each, curved storm types had three, and eye storm types had five.

Thus, the total number of questions per storm image (required plus optional questions) ranges from five to eight. The four storm types have different numbers of optional questions because different features are of interest to scientists based on the storm type. Storms with the fewest number of optional questions are most divisible, and those with the highest number of optional questions are least divisible in Version 1 of Cyclone Center. We refer to edits for which both required and optional questions were answered as *complete* edits, and those for which only the required questions are answered as *incomplete* edits. Figure 2 shows the required and optional questions in Version 1 for 'eye' storm types, those with the most optional questions (5).

Insert Figure 2 about here

Importantly, the storm type an editor works on is effectively random from the viewpoint of the editor. This is because editors are randomly shown a storm image, so they cannot choose the type to work on - rather, it is their task to identify it. Moreover, since Cyclone Center's goal is to track the behavior of storms throughout their lifetime, we argue that some storm types are not more interesting for editors to analyze than others, since every instance of a storm is important for mapping its behavior. This is important because it is a key identifying assumption in Part II of our empirical approach, described in Section 4.2.

The variation in the number of optional questions per storm image is a key feature that we exploit in the second part of our empirical approach, in which we use an additional empirical strategy to further test our theoretical predictions. It allows us to identify the relationship between contribution divisibility (i.e. operationalized through the number of optional questions) and editor completion rates of these questions. In Section 4.2, we use

this variation to compare the completion rates of optional questions for different storm types. Our model predicts a positive relationship between contribution divisibility and the total number of edits per editor. In the second part of our empirical approach, we further test this prediction by examining whether less divisible storm types (those with more optional questions) have lower completion rates than more divisible storm types.

3.2   Data

Our data consist of all edits made by Zooniverse editors in two projects, Cyclone Center and Galaxy Zoo, between September 2012 and February 2014, the first 70 weeks of each project. Our primary project, as discussed in Section 3.1, is Cyclone Center. We use Galaxy Zoo as a control group in Part I of our empirical approach, the difference-in-differences strategy, discussed in Section 4.1. Table 1 provides descriptive statistics on the two projects.

Insert Table 1 about here

In both projects, editors are asked to edit image files by answering questions via a decision-tree menu-based interface. We observe all edits that are submitted for each project. For each edit, we have the following information: the image ID, the date of the edit, the editor type (i.e. whether the editor is registered or anonymous), and the answers provided in each analysis. Over the 15-month observation period, there were 350,561 total edits made by editors in Cyclone Center. In Version 1 (first 36 weeks), there are 210,453 edits, while in Version 2 (34 weeks after the format change), there are 140,103 edits. Weekly, there are 10,800 total edits, 1,131 editors, and 9.5 edits per editor. In addition, storm images in Cyclone Center receive an average of 14.58 edits.

The second project, Galaxy Zoo, has 15,127,555 total edits, 7,143,699 in the time period that corresponds to Version 1 in Cyclone Center, and 7,983,856 in the period corresponding

to Version 2. Weekly, there are 203,141 total edits, 4,467 editors, and 45.47 edits per editor. Galaxy images receive an average of 19.63 edits. There are several reasons why Galaxy Zoo is an appropriate control project for the difference-in-differences strategy in Section 4.1. The project has a similar interface and question structure as Cyclone Center, but did not undergo any format changes during our observation period. Moreover, both projects ask editors to analyze images through visual inspection and the application of pattern recognition skills, relying on human ability to identify patterns in noisy images. Indeed, some of the web interface code from Galaxy Zoo was repurposed for Cyclone Center due to the similarities between the two projects' structures (Hennon et al. 2015). Figure 3 shows a sample of the types of images editors are asked to analyze in Cyclone Center and Galaxy Zoo, respectively.

Insert Figure 3 about here

The final sample consists of 15,478,116 edits, 422,172 images and 280,001 editors across the two projects.

4   Empirical Approach and Results

The main testable predictions of our theoretical model are that (i) an editor's probability of contributing total edits will be higher in a divisible than in a non-divisible editing format; (ii) it is ambiguous whether an editor's probability of contributing complete edits will be higher or lower in a divisible or a non-divisible format; and (iii) it is ambiguous whether completed contribution value will be higher or lower in a divisible or a non-divisible format. We employ a two-part empirical approach to test these predictions, where each part is based on a different set of assumptions.

In Part I of our empirical approach, we employ a difference-in-differences strategy, exploiting the Cyclone Center format change that lowered contribution divisibility by making the optional questions from Version 1 required. We present this strategy in Section 4.1. Part II of our empirical approach exploits the variation in contribution divisibility across different storm types in Version 1 of Cyclone Center. Recall that the number of optional questions in Version 1 varied depending on the storm type identified in the image. Fewer optional questions suggest higher contribution divisibility, while more optional questions imply lower contribution divisibility. Part II, presented in Section 4.2, exploits this variation to provide an additional empirical strategy for testing our theoretical predictions.

4.1 Part I: Difference-in-differences, exploiting Cyclone Center format change

Part I of our empirical approach employs a difference-in-differences strategy to investigate the relationship between contribution divisibility and editor contribution levels. To do so, we exploit the format change that occurred in Cyclone Center nine months after the project's launch. As discussed in Section 3.1, the format change decreased contribution divisibility by bundling together two tasks from Version 1 (required and optional questions). The format of Version 1 corresponds to a more divisible format, while the format of Version 2 corresponds to a more non-divisible one. Our theoretical model predicts that the number of total edits per editor will be higher in the divisible than the non-divisible format, and that it is ambiguous whether the number of complete edits and the complete edit value will be higher or lower in a divisible or a non-divisible format.

In an ideal experiment, we would test our theoretical predictions by (i) randomly assigning storm types with high contribution divisibility to one group of editors, and storm types with low divisibility to another, and (ii) comparing contributions between the two groups.

Although it was not possible to implement such an experiment, the format change in Cyclone Center simulates this ideal, because it represents a shift from a format with higher contribution divisibility to one with lower divisibility. Importantly, the change was unanticipated by Cyclone Center editors, as staff did not announce their plans for the change before its implementation.

Thus, the format change allows us to compare editor contributions across two formats with varying divisibility. However, simply comparing contribution levels before and after the format change is not enough to disentangle the effect of divisibility on contribution levels. Such a comparison would likely produce biased estimates, since we are unable to observe the counterfactual: editor contribution levels in the absence of a change. We perform this regression as shown in Equation (1):

$$y_{it} = \alpha + \beta_1 \times Post + \epsilon_{it} \tag{1}$$

where y_{it} is (i) the number of total edits made by editor i in week t, and (ii) the number of complete edits. Table 2 presents the results.

Insert Table 2 about here

We can see that there is a negative, significant correlation between the decrease in contribution divisibility resulting from the format change and total edits, and no significant relationship with the number of complete edits. However, this simple OLS regression of contribution levels on a pre/post format change indicator variable is likely to confound the effect of the change with possible temporal effects (i.e. time trends in contributions), due to the absence of the counterfactual. To properly isolate the effect of the change on editor contribution levels, we require a control group that would simulate the counterfactual, al-

lowing us to obtain the average change in contribution levels in the absence of a change. As described in Section 3.2, the project Galaxy Zoo, in which editors analyze galaxy images, provides an appropriate control group due to its similarities to Cyclone Center and lack of format changes.

We employ a difference-in-differences (DD) strategy, in which we compare the average change in contribution levels of Cyclone Center editors before/after the format change, to the average change in contribution levels of Galaxy Zoo editors before/after the time of the format change. In DD terminology, we refer to Cyclone Center as the treated project, Galaxy Zoo as the control project, and the Cyclone Center format change as the treatment.

The validity of the DD strategy rests on the critical assumption that in the absence of treatment, the average change in editor contribution levels would have been the same in both the treated and control projects. This is commonly referred to as the parallel trends assumption, because it requires the trends in contribution levels in the two projects to be similar prior to the time of treatment. We check this assumption in Section 4.1.3.

4.1.1 Methodology

Our main estimating equation, which examines the effect of a decrease in contribution divisibility (i.e. format change) on editor contribution levels, appears in Equation (2) below. The level of analysis is editor-week. We perform our analysis on the time period of four months before and after the format change, which occurred in week 36; this corresponds to weeks 20-35 for the pre-change period, and weeks 37-52 for the post-change period.

$$y_{it} = \alpha + \beta_1 \times CycloneCenter_i + \beta_2 \times (CycloneCenter_i \times Post_t) + \delta_t + \epsilon_{it} \qquad (2)$$

where y_{it} is (i) the number of total edits made by editor i in week t, and (ii) the number of complete edits. $CycloneCenter_i$ is a dummy variable equal to 1 for edits made by editors

in the treated project, Cyclone Center, to capture all time-invariant differences between the two projects. The main coefficient, β_2, captures the marginal effect of the format change (i.e. the decrease in contribution divisibility) on editor contribution levels. When the dependent variable is (i) the number of total edits, β_2 will be negative and statistically significant if editors respond to the format change by contributing fewer edits. When the dependent variable is (ii) the number of complete edits, β_2 will be negative if editors do fewer complete edits, and positive if editors do more complete edits. Week fixed effects, δ_t, are included to capture any changes over time that affect both the treated and control projects similarly. Robust standard errors are clustered by editor, in order to not overstate statistical significance due to possible serial correlation within editor (Bertrand et al. 2004).

4.1.2 Main Results

Table 3 presents the main results of the DD strategy. We have two main dependent variables: the number of total weekly edits per editor, and the number of complete weekly edits per editor, because we are interested in testing our model's theoretical predictions about both total and complete[3] edits following the Cyclone Center format change. Table 3 Model (1) shows the results for total weekly edits per editor. The coefficient of interest, on the variable $CycloneCenter \times Post$, is negative and statistically significant (-11.972); it suggests that editors in Cyclone Center respond to the decrease in contribution divisibility brought on by the format change by contributing roughly 12 fewer edits per week. Compared to an average of 24 weekly edits before the format change, this result corresponds to a 48% decrease in weekly edits after the format change. Table 3 Model (2) shows the results for complete weekly edits per editor. The coefficient of interest is positive, but insignificant, and the magnitude is small. This suggests that there is no significant change in the number of complete weekly edits per editor as a result of the format change.

---

[3] Recall that complete edits in Version 1 are those for which both the required and optional questions are answered.

After obtaining the baseline results, we run regressions of the form in Equation (2) for registered and anonymous editors separately. This is motivated by our theoretical model, which, in addition to the main predictions, further predicts that registered and anonymous editors may respond differently to the decrease in contribution divisibility resulting from the format change. This is because, assuming that registered editors have more experience with the platform than anonymous editors, they may constitute a class of editors with lower own contribution costs for the second task (lower a); the model predicts that such editors would be likely to contribute complete edits even when the format is non-divisible. Thus, we would expect their rate of contributing complete edits to be unchanged between the two Cyclone Center formats - the divisible format pre-change, and the non-divisible format post-change.

We empirically test whether the two types of editors respond differently to the decrease in contribution divisibility, and present the results in Table 4. First, Model (1) and Model (2) show the results for total weekly edits per editor, for registered and anonymous editors, respectively. The coefficient of interest, on the variable $CycloneCenter \times Post$, is negative and statistically significant for both editor types, suggesting that both editor types contribute fewer *total* edits weekly after the format change. Registered editors contribute 21.017 fewer edits weekly, while anonymous editors contribute 3.423 fewer edits. When comparing the size of these effects between the two editor types, it is important to note that the average number of total weekly edits is 43.628 for registered editors, and 7.382 for anonymous editors. This corresponds to a 64% reduction in total weekly edits for registered editors, and a 46% reduction for anonymous editors following the decrease in contribution divisibility resulting from the Cyclone Center format change.

23

Table 4 Models (3) and (4) present the results for *complete* weekly edits per editor, for registered and anonymous editors, respectively. For registered editors, the coefficient of interest is positive but insignificant, suggesting the format change has no significant effect on the number of complete weekly edits made by registered editors. However, for anonymous editors, the coefficient of interest is positive and significant, indicating that anonymous editors respond to the change by increasing the number of complete edits they make by 2.648 edits per week; this represents a large increase compared to the pre-change weekly average of 0.069 complete edits. These findings support our theoretical prediction that the rate of contributing complete edits would be unchanged for registered editors (but changed for anonymous editors) as Cyclone Center moves from a divisible to a non-divisible format, because relative to anonymous editors, registered editors have lower costs of contributing the second task (i.e. the questions that became required post-format change).

Insert Table 4 about here

4.1.3 Testing the 'Parallel Trends' Assumption

The key underlying assumption of the difference-in-differences strategy, discussed in Section 4.1, is that in the absence of treatment (i.e. the Cyclone Center format change), the average change in editor contribution levels would have been the same in both the treated (Cyclone Center) and control (Galaxy Zoo) projects. This assumption is commonly called the 'parallel trends assumption', because it requires similar trends in contribution levels in both projects prior to the format change. We test this assumption by formally examining the differences in editor contribution levels between Cyclone Center and Galaxy Zoo prior to the format change in Cyclone Center. To do so, we perform a regression that is similar to the main specification in Equation (2), but in which we replace the interaction term

24

CycloneCenter × *Post* with a series of dummy variables representing each week, interacted with the Cyclone Center indicator variable, as follows:

$$y_{it} = \alpha + \beta_1 \times CycloneCenter_i + \beta_2(CycloneCenter_i \times Week_t) + \delta_t + \epsilon_{it} \qquad (3)$$

This procedure allows us to check whether the trends in editor contribution levels are similar between the two projects prior to the format change in Cyclone Center. Figure 4 presents our results, for the time period of our analysis - weekly for four months before/four months after the format change. In Figure 4(a), the dependent variable is total weekly edits per editor, and each point on the graph represents the value of the coefficient on the variable *CycloneCenter_i* × *Week_t*. Thus, it captures the estimated difference in total weekly edits per editor between the treated and control projects in week *t*. Figure 4(b) shows this weekly difference for complete edits. The bars surrounding each point represent the 90 percent confidence intervals. The pre-treatment differences are not significantly different from zero in either Figure 4(a) or Figure 4(b).

This suggests that there are no significant differences in the number of total or complete weekly edits per editor between the treated and control projects prior to time of the format change. After the change in Cyclone Center, labeled as time 0 in both figures, the difference in contribution levels between projects becomes significant in Figure 4(a), but not in Figure 4(b). These findings are consistent with our regression results. The key takeaway from Figure 4 is that there are no significant differences in the pre-treatment trends in contribution levels between the two projects that may be driving the results, providing support to the key underlying 'parallel trends' assumption of the DD strategy.

Insert Figure 4(a)-(b) about here

4.1.4 Robustness Checks

Check 1 Our first main finding was that editors responded to the decrease in contribution divisibility brought on by the Cyclone Center format change by reducing their weekly number of total edits. This suggests that editors contributed less to the project as a result of the change. A possible question about this finding is whether editors are indeed doing less on the platform, or whether they still spend the same amount of time contributing, but are simply able to do fewer edits in that time because the change increased the number of questions they are required to answer per image. In order to address this question, we obtain data on the amount of time editors spend contributing to Cyclone Center, before and after the format change. The data are aggregated at the editor-version level. Recall that Version 1 is the pre-format change period, while Version 2 is the period post-change period. The data include the following variables: edit length, session length[4], number of edits per session, and total time spent contributing to Cyclone Center. In addition, for each editor we know the date of their first and last edit made in Cyclone Center.

Initial descriptive statistics in Table 5 show that not only do Cyclone Center editors make fewer edits post-format change, they also spend less time contributing overall. Moreover, post-change, the average time spent per edit nearly doubles due to the increase in the number of questions editors have to answer (time spent per edit increases from 1.27 minutes to 2.41 minutes per edit). At the same time, the session length decreases post-change, from 7.67 minutes to 6.64 minutes per session, corresponding to a 13% decrease. The total time editors spend contributing also decreases post-change, from 18.08 to 14.63 minutes, a 19% decrease.

Insert Table 5 about here

In order to further explore whether the changes in contributing time are due to the

---

[4] Session length is defined as the period of time between an editor entering and exiting Cyclone Center.

Cyclone Center format change, we obtain the same data for editors contributing to Galaxy Zoo, the control project. These data are also at the editor-version level, aggregated into two time periods, corresponding to the pre/post-format change periods in Cyclone Center. We perform a DD estimation to compare how the amount of time spent contributing changes after the time of the format change in Cyclone Center relative to Galaxy Zoo. We perform the following regression:

$$y_{it} = \alpha + \beta_1 \times CycloneCenter_i + \beta_2 \times Post_t + \beta_3(CycloneCenter_i \times Post_t) + \epsilon_{it} \qquad (4)$$

where y_{it} is (i) edit length, (ii) session length, (iii) number of edits per session, and (iv) total time spent contributing, respectively. Similar to our main specification, $CycloneCenter_i$ is a dummy variable equal to 1 for edits made by editors in the treated project, Cyclone Center, and $Post$ is a dummy variable equal to 1 for edits made after the time of the Cyclone Center format change. Robust standard errors are clustered by editor.

Table 6 presents our results. Table 6(a) presents the baseline result of the DD estimation, which explores the effect of the decrease in contribution divisibility due to the format change on contributing time per editor. Tables 6(b) and 6(c) separate this effect by editor type. Table 6(b) shows the effect for *registered* editors, and Table 6(c) for *anonymous* editors. Our findings show that anonymous editors respond to the decrease in contribution divisibility by not only reducing the number of edits they contribute, but also decreasing the time spent contributing on Cyclone Center - both session length and total time spent contributing by anonymous editors decrease post-change. Registered editors, on the other hand, do not significantly change the time spent contributing to the project. This finding provides additional evidence that registered and anonymous editors respond differently to the decrease in contribution divisibility, which we also found in our main results.

Note that a limitation of our approach in this subsection is that we are unable to formally test the DD 'parallel trends' assumption due to data constraints. Since our data are aggregated at the editor-version level (rather than the editor-week level as in our main specification), we only have two periods, one before and one after the format change; thus, we cannot observe the pre-trends. Due to this limitation, the results in this subsection are best interpreted as providing suggestive evidence of the relationship between contribution divisibility and contributing time per editor.

Check 2 A second possible question regarding our main findings is whether post-change, editors are more likely to skip analyzing images of storm types for which divisibility decreased the most. Recall that storms can be one of four possible types: embedded, shear, curved and eye. After the format change, the total number of required questions increased as follows: from 3 to 5 questions for embedded and shear storm types, from 3 to 6 questions for curved storm types, and from 3 to 8 questions for eye storm types.

A concern could be that editors may skip analyzing eye storm types (whose divisibility decreased the most), choosing instead to analyze images of the other three storm types. In order to investigate this possibility, we would ideally obtain data on the number of images of each storm type that are shown to editors before and after the format change, and the number of those images that they choose to analyze or abandon. However, we only observe analyzed edits (i.e. those that are submitted), not abandoned ones. Despite this data constraint, we do observe the number of edits of each storm type per editor. Thus, we perform the following regression to observe how the number of edits per editor changed for each storm type after the format change:

$$y_{sit} = \alpha + \beta_1 \times Post_t + \epsilon_{ist} \tag{5}$$

where y_{sit} is the number of total edits of storm type s made by editor i in period t. $Post$ is a dummy variable equal to 1 for edits made after the time corresponding to the format change. We run this regression for each of the four storm types: embedded, shear, curved and eye. Same as in our main specification, the pre-format change period is weeks 20-35, and the post-change period is weeks 37-52.

Table 7 presents our findings. We can see in Panel A (All Editors) that compared to the pre-change period, post-format change, the total number of embedded storm edits per editor decreased by 46%, shear storm edits decreased by 55%, curved storm edits decreased by 40%, and eye storm edits decreased by only 7%. When we further break down the results by editor type in Panel B (Registered Editors) and Panel C (Anonymous Editors), we can see that for registered editors, embedded storm edits decreased by 45%, shear storm edits decreased by 51%, curved storm edits decreased by 38%, and eye storm edits decreased by 2%. Regarding the anonymous editors, the number of embedded storm edits decreased by 75%, shear storm edits decreased by 70%, curved storm edits decreased by 75%, and eye storm edits decreased by 63%. Thus, we find that while the number of total edits per editor decreased for each storm type, the reduction in eye storm edits (for which contribution divisibility decreased the most) was actually the smallest among storm types. This result provides evidence against the concern that editors are more likely to skip analyzing eye storms (for which contribution divisibility decreased the most) than other storm types post-format change. Note that for the analysis in this subsection, we do not have an explicit control group for comparison, so the findings can be interpreted as providing suggestive evidence.

Insert Table 7 about here

4.1.5 Testing model assumption that editors care about total contributions

An assumption in our theoretical model is that editors care about overall edits on Cyclone Center, and not about their specific edits (i.e. public vs. private benefit). This assumption gives rise to the public good effects in the model. We test this assumption by examining whether the size of the crowd matters for a given editor's contribution levels. For instance, is there evidence of free-riding, where each editor contributes less when there are more editors on the platform? We have presented our model as a public good contribution model in order to provide a more general context, but the main theoretical predictions are robust to alternative specifications in case editors incur purely private benefits from contributing. It is a difficult econometric problem to test whether the size of the crowd affects editor contribution levels. Doing so would require an exogenous change in the crowd size, which was not possible to find. We begin to explore this relationship by running a regression similar to the main specification in Equation (2), adding the number of weekly editors as an explanatory variable.

$$y_{it} = \alpha + \beta_1 \times CycloneCenter_i + \beta_2 \times (CycloneCenter_i \times Post_t) + \beta_3 Editors_{it} + \delta_t + \epsilon_{it} \quad (6)$$

We present our findings in Table 8, which suggest that there is no significant relationship between the weekly number of editors and contribution levels per editor. However, it is important to note that this is merely a correlation, because there exist two main threats to establishing causality. First, the Cyclone Center format change may be a common factor impacting both the number of weekly edits per editor and the number of weekly editors. Second, there is a concern about reverse causality, since the number of weekly edits per editor may affect the number of editors, as follows: if editors contribute more edits, this could signal that the platform is becoming more popular and more editors may join due to word-of-mouth.

Due to these possible endogeneity concerns, further work is needed to disentangle the relationship between the size of the crowd and edits per editor in the context of crowdsourced science platforms. Previous work by Zhang and Zhu (2011) shows that an exogenous reduction in crowd size at Chinese Wikipedia leads the remaining volunteers to contribute less. The authors attribute this effect to the social gains accrued by contributors from belonging to the Wikipedia community, which shrink as the size of the community decreases.

Whether this relationship also holds in the context of crowdsourced science is an open question, and worthy of further investigation. This is because there is an important difference between crowdsourced science platforms and other crowdsourcing communities like Wikipedia: it takes longer to obtain the output of volunteer contributions on crowdsourced science platforms than on sites like Wikipedia. While the output of collaborative editing becomes instantly available to Wikipedia volunteers once they submit their edits, it takes a much longer time for crowdsourced science edits to lead to new scientific discoveries. Therefore, the most immediate benefits of crowdsouced science contributions accrue to the scientists conducting research, and only later to volunteers in the form of scientific discoveries, unlike in communities like Wikipedia.

There are several possible ways to disentangle the causality of the relationship between crowd size and editor contribution levels in Zooniverse. First, in recent months, Zooniverse has started translating its platform into multiple languages. In the future, this could potentially increase the number of editors. Second, some high school teachers have started using Zooniverse for classroom exercises, to give their students the opportunity to participate in real world scientific research. By tracking the locations of the IP addresses, it may be possible to figure out which ones are associated with high school or middle schools. This could

31

provide an exogenous increase in the number of editors on the platform when a class begins and students start contributing.

4.2 Part II: Exploiting variation in contribution divisibility prior to Cyclone Center format change

In Part II of our empirical approach, we present an additional, alternative strategy to test our second theoretical prediction: whether an editor's probability of contributing complete edits will be higher or lower in a divisible or a non-divisible format. For this strategy, we exploit the variation in contribution divisibility across different storm types in Version 1 of Cyclone Center. Recall that during Version 1, which lasted for nine months until the format change, each storm image had three required questions, and different numbers of optional questions. The number of optional questions depended on the storm type identified by an editor in one of the required questions. A storm's type indicates its intensity at a given time, and varies over the course of the storm. It can be categorized into four possible types: embedded, shear, curved and eye. Embedded and shear storm types had two optional questions, curved storm types had three, and eye storm types had five.

We capture contribution divisibility through the number of optional questions per storm type, where a lower number of optional questions implies higher contribution divisibility. We test whether an editor's probability of contributing complete edits - edits for which both required and optional questions are answered - is higher or lower for more divisible storm types (e.g. embedded and shear, with the lowest number of optional questions) than for non-divisible storm types (e.g. eye, with the highest number of optional questions).

The main identifying assumption in this part of our empirical approach is that some storm types are not more interesting for editors to analyze than others (i.e. that the storm type in an image does not affect editors' propensity to analyze it). We argue that this is a

reasonable assumption because the goal of Cyclone Center is to track the behavior of storms throughout their lifetime. Thus, each image represents an important component in tracking the behavior of a storm. A second important assumption is that the storm type in a given image is effectively random from the point of view of the editor. We argue this is reasonable because neither editors nor Cyclone Center staff know a storm's type before analyzing the image, since storm type (which captures intensity) varies throughout the course of a storm.

We run a regression of whether or not an edit is complete in Version 1 based on the storm type. We run a regression of the following form:

$$Complete_i = \alpha + \beta_1 \times Embedded_i + \beta_2 \times Shear_i + \beta_3 \times Curved_i + \beta_4 \times Eye_i + \epsilon_i \quad (7)$$

where $Complete_i$ is an indicator variable that captures whether or not edit i is complete or incomplete; $Embedded_i$, $Shear_i$, $Curved_i$ and Eye_i are indicator variables to reflect the storm type in image i from the four possible storm types.

Table 9 presents the findings, showing the relationship between the probability of completeness and the storm type. The baseline for comparison in the regression is the eye storm type. Our model predicts that storm types with higher contribution divisibility (i.e. fewer optional questions - embedded, shear and curved storm types) would have a higher probability of completeness than the storm type with the lowest contribution divisibility (i.e. highest number of optional questions - eye storm type). We can see in Table 9 that there is no clear relationship between contribution divisibility and probability of completeness. Compared to eye storms, only embedded storms have a significantly higher probability of being completed.

Insert Table 9 about here

Next, we divide the data by editor type, to examine whether registered and anonymous

33

editors respond differently to the differences in contribution divisibility across storm types. Recall that we also tested this theoretical prediction through the difference-in-differences approach in Part I, and found evidence that for registered editors, their rate of contributing complete edits was unchanged before and after the format change (in both divisible and non-divisible formats), while for anonymous editors, their rate of contributing complete edits was higher post-change, in the non-divisible format.

We run the same regression as in Equation (7) for each editor type separately. Table 10 presents the findings. We can see that for registered editors, similar to the main findings in Table 9, there is no clear relationship between contribution divisibility (operationalized through storm type) and probability of completeness. Turning to the anonymous editors, we can see that their probability of completing storm types with fewest optional questions (embedded and shear) is significantly higher than their probability of completing eye storm types, with the highest number of optional questions. These findings suggest that for anonymous editors, the higher the number of optional questions, the lower the probability of them answering these questions.

Insert Table 10 about here

4.3   The Value of *Complete* Edits

The third main theoretical prediction of our model is that it is ambiguous whether completed contribution value will be higher or lower in a divisible or a non-divisible format. Recall that complete edits are those in which an editor has answered both the required and optional questions about a storm image, while incomplete edits are those in which only the required questions have been answered. Two factors primarily contribute to the value of complete edits: (i) the importance of complete edits for scientists using the Cyclone Center

34

data, and (ii) the quality of complete edits. While it is difficult to determine how important complete edits are to Cyclone Center relative to incomplete edits, the fact that Cyclone Center implemented a format change whose aim was to obtain more complete edits (by making previously optional questions required) suggests that complete edits were valuable. Then, we turn to investigating the second factor: the quality of complete edits, which affects the value of complete contributions. Our findings in Section 4.1 showed that after the format change, anonymous editors contributed more *complete* edits. What is the implication of this result for the overall quality of complete edits?

We evaluate complete edit quality by measuring the accuracy of editors' answers to the optional questions (those which distinguished complete from incomplete edits). We proxy accuracy with editors' consensus in their answers to each optional question, which we operationalize through a Herfindahl-Hirschman Index (HHI). Similar consensus-based measures have been used in prior research (Hennon et al. 2015) to capture a crowd's performance in accurately identifying storm features in Cyclone Center. Recall that each storm image is shown until it obtains approximately ten edits, in order to improve the accuracy of the classification. We calculate a consensus measure for each optional question by editor type, to examine whether the increase in the number of complete edits contributed by anonymous editors after the format change has positive or negative implications for edit quality. We interpret a higher consensus measure as higher accuracy in answering a question.

For each optional question q in storm image i, we construct a consensus measure, operationalized through a HHI, as follows:

$$HHI_{q,i} = \sum_x s_{x,q,i}^2$$

where x is the set of possible answers for optional question q, and $\sum_x s_{x,q,i} = 1$ (i.e. the sum of the shares of each possible answer to q is 1).

Table 11 shows the consensus measure for each optional question, by editor type (reg-

istered or anonymous). Since each of the four possible storm types has different optional questions associated with it, the questions are presented by storm type. Findings show that the consensus among anonymous editors is higher for each optional question than the consensus among registered editors. Thus, if we interpret editors' consensus as a proxy for accuracy, our findings suggest that the value of complete edits after the format change (i.e. in the non-divisible format) is higher than before the format change (i.e. in the divisible format). This is because anonymous editors responded to the format change by increasing the number of complete edits they contributed, and they had high accuracy in answering the additional questions that made the edits complete.

Insert Table 11 about here

4.3.1 Strategic Implications

The analysis in this section suggests that the choice of contribution design - specifically, the divisibility of contributions - matters for crowdsourcing platforms. Relevant factors in this decision include: (i) the level of difficulty and/or specialized knowledge needed to contribute, (ii) the type of data the crowdsourcing platform seeks to collect, and (iii) the skill level of the volunteers.

For example, if small contributions provide useful data for a crowdsourcing platform, the knowledge required to contribute is basic, and the volunteers possess this knowledge, a divisible contribution format could be preferred. Alternatively, if useful contributions consist of substantial edits requiring more specialized knowledge, then a non-divisible format could be more suitable.

5 Conclusion

Technological advances in the past two decades have led to the rise of crowdsourcing, a collaborative form of content production based on the contributions of volunteers that typically takes place online (Howe 2006, Zhao and Zhu 2014). A well-known example of crowdsourcing is Wikipedia. Rather than employ writers to create an online encyclopedia, Wikipedia outsources its content production entirely to volunteers, who collaborate to write articles (Greenstein and Zhu 2012). In recent years, crowdsourcing has been adopted in a variety of areas, including business, government, and science. Due to this proliferation, scholars have turned their attention to understanding the factors that affect the sustainability of crowdsourcing initiatives (Zhao and Zhu 2014).

One factor that has received limited attention in the literature is how the design of crowdsourcing platforms affects volunteer contributions. In this paper, we have focused on one particular design feature - the divisibility of contributions (i.e. whether contributing tasks are bundled together or can be carried out selectively) - and studied whether this is a factor in the level and quality of crontributions on crowdsourcing platforms.

We investigated this question in the context of Zooniverse, the world's largest crowd-sourced science platform, specifically focusing on one of its projects, Cyclone Center. The objective of Cyclone Center is to help scientists understand the behavior of tropical storms in order to better predict future storms; volunteers contribute by analyzing satellite images of storms. Analysis initially consisted of two tasks: (1) answering three required questions, and (2) answering several additional optional questions. We refer to contributions in which both tasks are done as *complete*, and contributions in which only the first task is done as *incomplete*.

Two features of Cyclone Center made this project an appropriate empirical setting. The first is a format change that occurred in Cyclone Center nine months after the project's

launch. The format change *decreased* contribution divisibility, by bundling together the two tasks that could previously be done selectively. We exploited this change in the first part of our empirical approach, employing a difference-in-differences strategy. The second feature that made Cyclone Center a suitable setting is the structure of the project prior to the format change. During this time, there was variation in contribution divisibility across storm images, because the number of optional questions varied based on the storm type identified by volunteers in a required question. We used this variation in the second part of our empirical approach, to further explore the relationship between contribution divisibility and volunteer contributions.

In the context of Zooniverse, we developed a theoretical model and tested its theoretical predictions through our two-part empirical approach. We found that when the contribution design in Cyclone Center changed from a divisible to a non-divisible one, (i) the total number of contributions per volunteer decreased, (ii) the number of complete contributions made by anonymous volunteers increased, and (iii) the value of complete contributions was higher because anonymous volunteers made high quality contributions. Our results have strategic implications for crowdsourcing platforms because they suggest that contribution design, and particularly the divisibility of contributions, matters for the quality and level of volunteer contributions. When making this decision, several factors are important to consider, including the type of contributions that are most valuable for the platform, the necessary expertise to contribute, and the skill set of the volunteers.

References

[1] Aaltonen, A. and S. Seiler. (2016). "Cumulative Growth in User-Generated Content Production: Evidence from Wikipedia," *Management Science*, 62(7): 2054-2069.

[2] Bayus, B. L. (2013). "Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community," *Management Science*, 59(1): 226-244.

[3] Bertrand, B., Duflo , E., and S. Mullainathan. (2003). "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics*, 1-32.

[4] Boudreau, K. and K. R. Lakhani. (2013). "Using the crowd as an innovation partner," *Harvard Business Review*, 61-69.

[5] Dvorak, V. F. (1973)."A Technique For the Analysis and Forecasting of Tropical Cyclone Intensities From Satellite Pictures," *National Oceanic and Atmospheric Administration*, 5-8.

[6] Gallus, J. (2016). "Fostering Public Good Contributions with Symbolic Awards: A Large-Scale Natural Field Experiment at Wikipedia," *Management Science*, Articles in Advance: 1-17.

[7] Greenstein, S. and F. Zhu. (2017). "Do Experts or Crowd-Based Models Produce More Bias? Evidence from Encyclopædia Britannica and Wikipedia", Forthcoming, *Management Information Systems Quarterly*.

[8] Halfaker, A., Geiger, R. S., Morgan, J. T., and J. Riedl. (2013). "The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline," *American Behavioral Scientist*, 57(5), 664-688.

[9] Hennon, C.C., Knapp, K.R., Schreck III, C.J., Stevens, S.E., Kossin, J.P., Thorne, P.W., Hennon, P.A., Kruk, M.C., Rennie, J., Gadea, J.M., and M. Striegl. (2015). "Cyclone Center: Can Citizen Scientists Improve Tropical Cyclone Intensity Records?" *Bulletin of the American Meteorological Society*, 96(4): 591-607.

[10] Howe, J. (June 1 2006). "The rise of crowdsourcing," *Wired*. Available at: http://www.wired.com/wired/archive/14.06/crowds.html.

[11] Huang, Y., P. V. Singh, and K. Srinivasan. (2014). "Crowdsourcing New Product Ideas Under Consumer Learning," *Management Science*, 60(9): 2138-2159.

[12] Lakhani, K. R., Jeppesen, L. B., Lohse, P. A., and J. A. Panetta. (2007). "The value of openess in scientific problem solving," Harvard Business School Working Paper, No. 07-050.

[13] Nov, O. (2007). "What Motivates Wikipedians?," *Communications of the ACM*, 50(11): 60-64.

[14] Nov, O., Arazy, O., and D. Anderson. (2011). "Technology-Mediated Citizen Science Participation: A Motivational Model," *Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM 2011).* Barcelona, Spain, July 2011.

[15] Peddibhotla, N. B., and M. R. Subramani. (2007). "Contributing to public document repositories: A critical mass theory perspective," *Organization Studies*, 28(3): 327-346.

[16] Rashid, A. M., Ling, K, Tassone, R.D., Resnick, P., Kraut, R., and J. Riedl. (2006). "Motivating participation by displaying the value of contribution," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI 2006, Montreal, Canada, 955-958.

[17] Schroer, J., and G. Hertel. (2009). "Voluntary Engagement in an Open Web-Based Encyclopedia: Wikipedians and Why They Do It" *Media Psychology*, 12: 96-120.

[18] See L., Comber A., Salk C., Fritz S., van der Velde M., Perger, C., Schill, C., McCallum, I., Kraxner, F., and M. Obersteiner. (2013). "Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts," *PLoS ONE*, 8(7): e69958. https://doi.org/10.1371/journal.pone.0069958

[19] Segal, A., Gal, Y., Kamar, E., Horvitz, E., Bowyer, A., and G. Miller. (2016). "Intervention Strategies for Increasing Engagement in Crowdsourcing: Platform, Predictions, and Experiments," *Proceedings of the 25th International Joint Conference on Artificial Intelligence.*

[20] Simpson, R., Page, K.R. and D. De Roure. (2014). "Zooniverse: observing the world's largest citizen science platform," *Proceedings of the companion publication of the 23rd International Conference on World Wide Web*: 1049-1054. Available at: http://dx.doi.org/10.1145/2567948.2579215.

[21] Zhao, Y. and Q. Zhu. (2014). "Evaluation on crowdsourcing research: Current Status and future direction," *Information Systems Frontiers*, 16: 417-434.

[22] Zhang, X. and F. Zhu. (2011). "Group size and incentives to contribute: A natural experiment at Chinese Wikipedia." *The American Economic Review*, 101(4): 1601-1615. Harvard

Table 1: Descriptive Statistics for Cyclone Center and Galaxy Zoo

Project	Cyclone Center	Galaxy Zoo
Launch Date	September 2012	September 2012

Overall (70 weeks)

Total		
Edits	350,561	15,127,555
Editors	20,257	259,744
Weekly Average		
Edits	10,800	203,141
Editors	1,131	4,467
Edits per Editor	9.5	45.47

Version 1 (Weeks 1-36)

Total		
Edits	210,453	7,143,699
Editors	11,381	124,791
Weekly Average		
Edits	15,003	264,887
Editors	1,437	6,594
Edits per Editor	10.44	40.17

Version 2 (Weeks 37-70)

Total		
Edits	140,103	7,983,856
Editors	8,874	134,953
Weekly Average		
Edits	4,487	147,893
Editors	758	2,847
Edits per Editor	5.91	51.95

Table 1 presents descriptive statistics for Cyclone Center and Galaxy Zoo, the two projects used in Part I of the empirical analysis. The variables include the *total* and *weekly* number of edits and editors, as well as weekly edits per editor. The table presents total numbers for the entire time period under observation, as well as separately for the time periods before and after the week of the format change in Cyclone Center (week 36).

Table 2: Simple Correlation Between Contribution Divisibility *Decrease* and Contribution Levels

Variable	DV = *Total* weekly edits per editor	DV = *Complete* weekly edits per editor
	Model (1)	Model (2)
Post	-14.025***	-0.512
	(2.677)	(2.190)
Observations	5,190	5,190
Clusters	4,407	4,407
R^2	0.02	0.01

Table 2 presents the results of the simple OLS regression exploring the relationship between a decrease in contribution divisibility due to the Cyclone Center format change and editor contribution levels. The dependent variable in Model (1) is the number of *total* weekly edits per editor. In Model (2), it is the number of *complete* weekly edits per editor.

Robust standard errors clustered at the editor level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Effect of Contribution Divisibility *Decrease* (Format Change) on Contribution Levels

Variable	DV = *Total* weekly edits per editor	DV = *Complete* weekly edits per editor
	Model (1)	Model (2)
Cyclone Center × Post	-11.972***	2.209
	(3.207)	(2.143)
Cyclone Center	-1.541	9.270***
	(3.394)	(2.838)
Mean of DV in Cyclone Center (before the format change)	24.482	0.393
Observations	86,772	86,772
Clusters	71,395	71,395
R^2	0.01	0.06
Week FE (32 weeks)	YES	YES

Table 3 presents the results of the difference-in-differences estimation, showing the effect of the Cyclone Center format change, which led to a decrease in contribution divisibility, on editor contribution levels. The dependent variable in Model (1) is the number of *total* weekly edits per editor. In Model (2), it is the number of *complete* weekly edits per editor.

Regressions include week fixed effects for the period of analysis (4 months before and after the week of the format change: week 36). Robust standard errors clustered at the editor level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Means of DVs are also presented.

Table 4: Effect of Contribution Divisibility *Decrease* (Format Change) on Contribution Levels, by Editor Type

	DV = *Total* weekly edits per editor		DV = *Complete* weekly edits per editor	
Variable	Registered	Anonymous	Registered	Anonymous
	Model (1)	Model (2)	Model (3)	Model (4)
Cyclone Center × Post	-21.017***	-3.423***	3.241	2.648***
	(5.986)	(0.958)	(3.591)	(0.217)
Cyclone Center	-8.429***	-0.496	15.088***	1.084***
	(5.581)	(0.621)	(4.726)	(0.218)
Mean of DV in Cyclone Center (before the format change)	43.628	7.382	0.754	0.069
Observations	40,385	46,387	40,385	46,387
Clusters	28,276	43,119	28,276	43,119
R^2	0.01	0.01	0.09	0.12
Week FE (32 weeks)	YES	YES	YES	YES

Table 4 presents the results of the difference-in-differences estimation, showing the effect of the Cyclone Center format change, which led to a decrease in contribution divisibility, on editor contribution levels, for registered and anonymous editors separately. The dependent variable in Models (1) and (2) is the number of *total* weekly edits per editor, for registered and anonymous editors, respectively. The dependent variable in Models (3) and (4) is the number of *complete* weekly edits per editor, for registered and anonymous editors, respectively.

Regressions include week fixed effects for the period of analysis (4 months before and after the week of the format change: week 36). Robust standard errors clustered at the editor level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Means of DVs are also presented.

Table 5: Editing Time, Before and After Format Change

Variable	Before	After
All Editors		
Edit Length	1.27	2.41
Session Length	7.67	6.64
Total Time	18.08	14.63
Registered Editors		
Edit Length	1.21	2.40
Session Length	12.15	10.38
Total Time	35.87	29.78
Anonymous Editors		
Edit Length	1.31	2.41
Session Length	5.12	4.24
Total Time	6.35	4.92

Table 5 presents editors' average editing duration (in minutes) before and after the format change. Variables include edit length (minutes spent per storm image), session length (minutes spent from entry to exit on the platform), and total time (overall minutes spent by editors on platform).

Table 6(a): Effect of Contribution Divisibility *Decrease* (Format Change) on Editing Time

Variable	DV = Time/edit (minutes)	DV = Session length (minutes)	DV = Edits/session (number)	DV = Total time contributing (minutes)
Cyclone Center × Post	1.101***	-1.160**	0.024	-4.006
	(0.090)	(0.472)	(0.554)	(2.466)
Cyclone Center	0.339***	-0.217	-12.340***	0.279
	(0.047)	(0.400)	(0.442)	(1.569)
Post	0.041***	-0.811***	-3.475***	0.946
	(0.014)	(0.111)	(0.423)	(0.648)
Mean of DV in Cyclone Center	0.985	8.107	18.863	17.029
Observations	72,313	85,603	85,603	85,603
Clusters	71,778	84,908	84,908	84,908
R^2	0.01	0.01	0.01	0.01

Table 6(a) shows the effect of a decrease in contribution divisibility (i.e. format change) on editing time. Regressions are performed on edits made during the period of the main specification (edit ors who contributed in the period 4 months before and 4 months after the week of the format change: week 36). Robust standard errors clustered at the editor level are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6(b): Effect of Contribution Divisibility *Decrease* (Format Change) on Editing Time, for Registered Editors

Variable	DV = Time/edit (minutes)	DV = Session length (minutes)	DV = Edits/session (number)	DV = Total time contributing (minutes)
Cyclone Center × Post	1.195***	1.101	1.393	-8.719
	(0.105)	(0.707)	(1.291)	(6.194)
Cyclone Center	0.224***	-6.948***	-33.545***	-8.251**
	(0.042)	(0.502)	(0.978)	(3.885)
Post	0.062***	-0.794***	-5.102***	8.195***
	(0.021)	(0.259)	(1.137)	(1.935)
Mean of DV in Cyclone Center	0.933	16.204	40.970	39.357
Observations	26,698	28.272	28,272	28,272
Clusters	26,324	27,791	27,791	27,791
R^2	0.02	0.01	0.01	0.01

Table 6(b) shows the effect of a decrease in contribution divisibility (i.e. format change) on editing time for registered editors. Regressions are performed on edits made during the period of the main specification (editors who contributed in the period 4 months before and 4 months after the week of the format change: week 36). Robust standard errors clustered at the editor level are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6(c): Effect of Contribution Divisibility *Decrease* (Format Change) on Editing Time, for Anonymous Editors

Variable	DV = Time/edit (minutes)	DV = Session length (minutes)	DV = Edits/session (number)	DV = Total time contributing (minutes)
Cyclone Center × Post	1.080***	-3.677***	-3.266***	-5.672***
	(0.141)	(0.609)	(0.495)	(1.092)
Cyclone Center	0.413***	3.248***	-1.285***	4.133***
	(0.072)	(0.560)	(0.409)	(0.981)
Post	0.021	0.076	-0.232	0.051
	(0.018)	(0.091)	(0.252)	(0.325)
Mean of DV in Cyclone Center	1.016	4.115	7.961	6.018
Observations	45,615	57,331	57.331	57,331
Clusters	45,454	57,117	57,117	57,117
R^2	0.01	0.01	0.01	0.01

Table 6(c) shows the effect of a decrease in contribution divisibility (i.e. format change) on editing time for anonymous editors. Regressions are performed on edits made during the period of the main specification (editors who contributed in the period 4 months before and 4 months after the week of the format change: week 36). Robust standard errors clustered at the editor level are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7: Relationship Between Contribution Divisibility *Decrease* (Format Change) and Number of Edits of Each Storm Type

Variable	DV = Embedded Storm Edits	DV = Shear Storm Edits	DV = Curved Storm Edits	DV = Eye Storm Edits
Panel A: All Editors				
Post	-42.968***	-16.276***	-20.674***	-2.409***
	(0.705)	(0.560)	(0.448)	(0.315)
Mean of DV (before the format change)	90.293	29.375	50.611	32.766
Observations	81,758	81,758	81,758	81,758
R^2	0.04	0.01	0.02	0.01
Panel B: Registered Editors				
Post	-47.657***	-17.892***	-22.468***	-1.256***
	(0.800)	(0.661)	(0.513)	(0.358)
Mean of DV (before the format change)	103.238	32.850	57.706	36.822
Observations	68,711	68,711	68,711	68,711
R^2	0.04	0.01	0.02	0.01
Panel C: Anonymous Editors				
Post	-15.504***	-7.085***	-9.566***	-7.136***
	(0.519)	(0.311)	(0.310)	(0.349)
Mean of DV (before the format change)	20.855	10.733	12.556	11.015
Observations	13,047	13,047	13,047	13,047
R^2	0.06	0.03	0.06	0.03

Table 7 shows the correlations between the decrease in contribution divisibility due to the format change and the number of edits per editor of each storm type.

Table 8: Relationship Between Number of Editors and Contributions per Editor

Variable	DV = *Total* weekly edits per editor
Cyclone Center × Post	-13.989***
	(3.202)
Cyclone Center	-0.104
	(3.682)
Editors	0.001
	(0.001)
Observations	86,721
Clusters	71,361
R^2	0.01
Week FE (32 weeks)	YES

Table 8 shows the correlation between weekly number of editors and the average number of contributions per editor in Cyclone Center. Regressions include week fixed effects for the period of analysis (4 months before and after the week of the format change: week 36). Robust standard errors clustered at the editor level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9: Relationship Between Probability of Completeness and Storm Type Before Format Change

Storm Type	DV = Complete
Embedded	0.091***
	(0.006)
Shear	-0.152***
	(0.008)
Curved	0.009
	(0.007)
Observations	45,377
R^2	0.02

Table 9 shows a regression of whether or not an analysis is complete based on the storm type in Version 1 of Cyclone Center, before the format change. The results are shown as compared to the baseline (not included) type, the eye storm type, with the highest number of optional questions in Version 1.

Table 10: Relationship Between Probability of Completeness and Storm Type Before Format Change, by Editor Type

Registered Editors	
Storm Type	DV = Complete
Embedded	0.080***
	(0.007)
Shear	-0.201***
	(0.009)
Curved	-0.007
	(0.007)
Observations	38,395
R^2	0.03
Anonymous Editors	
Embedded	0.063***
	(0.014)
Shear	0.068***
	(0.017)
Curved	0.046
	(0.016)
Observations	6,982
R^2	0.02

Table 10 shows a regression of whether or not an analysis is complete based on the storm type in Version 1 of Cyclone Center, before the format change. The analysis is broken up by editor type. The results are shown as compared to the baseline (not included) type, the eye storm type, with the highest number of optional questions in Version 1.

Table 11: Consensus-Based Measure (HHI) for Each Optional Question

	Consensus Measure (HHI)	
	Registered Editor	Anonymous Editor
Embedded Storm Type		
Optional Questions (2)		
1. center(x,y)	0.207	0.521
2. banding feature	0.628	0.811
Shear Storm Type		
Optional Questions (2)		
1. center(x,y)	0.262	n/a
2. red point nearest (x,y)	0.892	0.907
Curved Storm Type		
Optional Questions (3)		
1. center(x,y)	0.306	n/a
2. band wrap	0.802	0.844
3. coldest band color	0.802	0.843
Eye Storm Type		
Optional Questions (5)		
1. center(x,y)	0.214	0.250
2. banding feature	0.781	n/a
3. coldest at least 0.5 degrees	0.899	0.927
4. coldest surrounding eye	0.899	0.927
5. eyewall	0.899	0.927

Table 11 presents editors' consensus when answering each optional question associated with a storm image (each image receives on average ten contributions). The focus is on optional questions because answering those questions (in addition to the required questions) is what makes an edit complete. Consensus is operationalized through an Herfindahl-Hirschman Index (HHI), and is computed separately for each optional question. In the table, the consensus measure is shown separately for each of the four different storm types (embedded, shear, curved, and eye). Recall that "optional questions" are those questions that were optional prior to the format change.

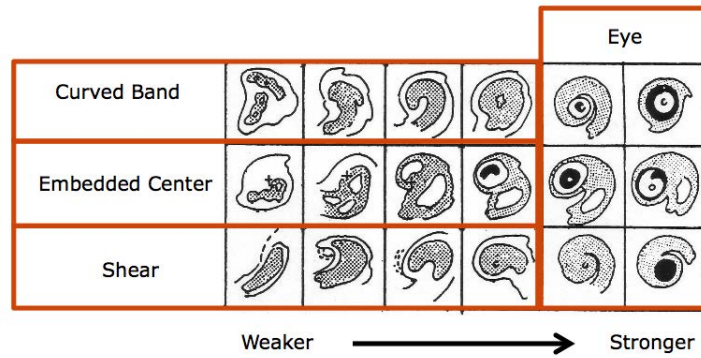# Figure 1: Four Storm Types Based on the Dvorak Technique

Figure 1 shows the four storm types identified by the Dvorak Technique, a widely used method among climatologists to determine the intensity of a storm based on its cloud patterns. The four storm types are: curved, embedded, shear, and eye.

Note that Cyclone Center editors are asked to identify a storm's type based on these four choices.
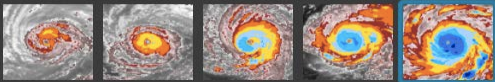
Figure 2: Question Workflow for Cyclone Center "Eye" Storm Types

Figure 2 shows the questions that Cyclone Center editors are asked to answer for "eye" storm types. The questions in the section "Three Required Questions" were required both before and after the Cyclone Center format change in order to submit an edit. The questions in the section "Five Additional Questions" were originally optional, and were made required after the format change.

Figure 3: Sample Images Shown In Cyclone Center and Galaxy Zoo

CYCLONE CENTER
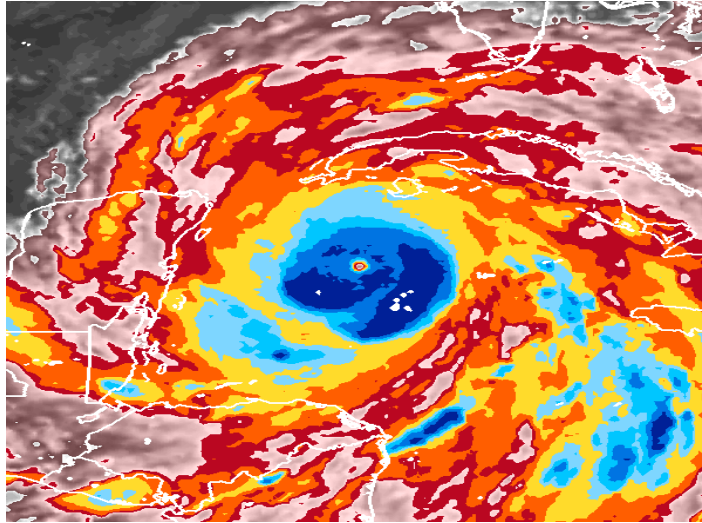
GALAXY ZOO

Figure 3 shows samples of the images that editors are asked to analyze in Cyclone Center and Galaxy Zoo, respectively.

Figure 4(a): Differences in Contribution Levels of *Total Edits* Between Cyclone Center & Galaxy Zoo, Before and After Cyclone Center Format Change
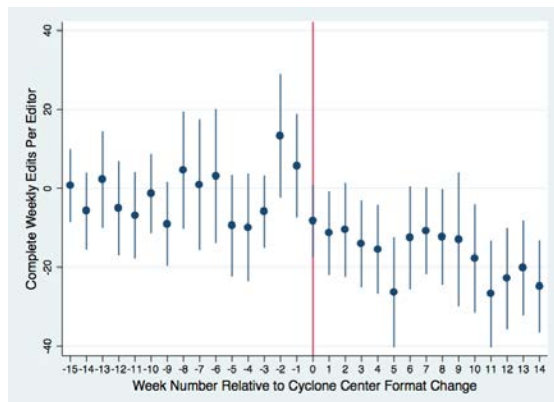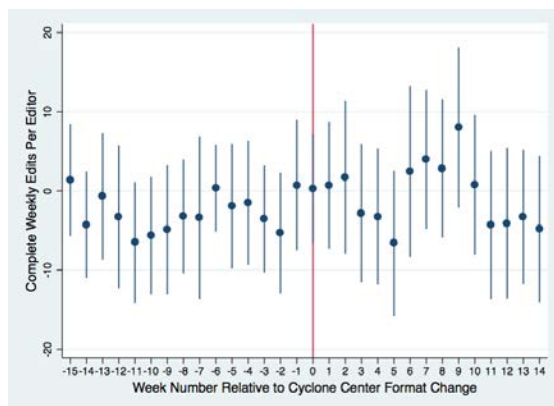

Figure 4(b): Differences in Contribution Levels of *Complete* Edits Between Cyclone Center & Galaxy Zoo, Before and After Cyclone Center Format Change

$$y_{it} = \alpha + \beta_1 \times CycloneCenter_i + \beta_2(CycloneCenter_i \times Week_t) + \delta_t + \epsilon_{it}$$

The graphs in Figure 4 are based on the point estimates and standard errors for β_2 in Equation (2), reproduced above. Each point represents the coefficient value on the variable $CycloneCenter \times Week$, and describes the relative difference in the weekly contribution levels of editors in the "treated" project, Cyclone Center, and those in the "control" project, Galaxy Zoo. The bars represent the 90 percent confidence intervals. In Figure 4(a), the dependent variable in the regression above is *total* weekly edits per editor. In Figure 4(b), it is *complete* weekly edits per editor.

Appendix A

Proof that when considering Task 2, the agent chooses to do the task with probability one so long as the high cost outcome does not arise.

Step One: Note that under the assumption that $c_H > V - v$, it is never optimal for the agent to choose to do the second task.

Step Two: For other cost realizations, the agent will choose to do the task with certainty if:

$$V - v - max[ac_h, c_L] > \delta(\frac{1}{n}\pi(1) + \frac{n-1}{n}\pi(0)) = \tilde{V} - \tilde{C}$$

where the last step follows when $q = \tilde{q}$ for the first task. This is true by assumption.

$$(V - v)(1 - \alpha)\gamma - v + c + a\alpha\gamma c_H + (1 - (1 - a)\alpha)(1 - \gamma)c_L > max[ac_H, c_L]$$