UNIVERSITAT DE BARCELONA

Estimating motion and time to contact in 3D environments: Priors matter

Borja Aguado Ramírez

UNIVERSITAT DE BARCELONA

Estimating motion and time-to-contact in 3D environments: Priors matter

Borja Aguado Ramírez

Department of Cognition, Development and Psychology of Education,

Universitat de Barcelona

Ph.D. Programme: Brain, Cognition and Behaviour

Supervisor: Joan López-Moliner

Barcelona, 03-2022

Contents

A mis padres, Fernando y Prado.

Agradecimientos

En nuestro día a día no solemos dedicar un espacio a dar las gracias. Creo que por eso - la falta de costumbre - ésta es la sección que más me está costando escribir.

Joan, gracias. Por todo. Recuerdo con cariño una tarde en la que me dijiste que trabajaba para mí mismo, no para ti ni para nadie de la universidad. Creo que eso me ayudó a entender cuál sería el rol de cada uno. En esta etapa tú me has proporcionado las herramientas y un entorno seguro en el que poder ser lo más autónomo posible, y eso tiene un valor incalculable. Por eso, gracias por cada una de las charlas en tu despacho o por Teams. Por cada vez que me ponía palos en las ruedas y tú me ayudabas a aclarar mis ideas. Por enseñarme a entusiasmarme con una idea, a hacerme *buenas* preguntas y a saber buscar por mí mismo las respuestas. Por estar siempre disponible para "un momentito" más. Por otra parte, y en un tono más personal, darte las gracias por el resto de las cosas que me has enseñado. Desde gastronomía lusa pasando por enología, apicultura, geografía catalana o Star Trek.

Cristina de la Malla, eres la persona del lab a quien más tiempo llevo conociendo. Desde mi Erasmus en Àmsterdam ya estuviste ahí para ayudar a aquel pollito que se iniciaba en la investigación. A tu llegada a Barcelona yo ya me encontraba haciendo el doctorado y, por supuesto, seguiste siendo un modelo para mí. Te doy las gracias por todas las charlas en el despacho, por los yogures de la tarde y por sobrellevar mi *a veces* excesiva expresividad.

Este es uno de los momentos más dulces. El momento de agradecer a aquellos que han hecho del despacho un chiquipark. Juntos hemos compartido momentos duros, hemos llorado de risa, hemos discutido, hemos llegado a la conclusión de que la monarquía neerlandesa es la mejor del

mundo, etc. Allí dejo un trocito de mi vida y un pequeño legado. El "hall de la fama". Continuadlo, o no. En tu mano queda Jaume ;). Vosotros, compañeros, habéis sido un apoyo fundamental en ese viaje y ahora es mi momento de dedicaros unas palabras: David, no compartimos mucho tiempo juntos, pero sin duda me quedo con tu oscuro sentido del humor. Björn, mi futuro IP cuando se colonice Marte y ejemplo a seguir. Genio de la cocina improvisada, militante por la movilidad sostenible y la identidad personal. He aprendido tantísimo de tu orgullo y de tu sencillez... Espero que podamos colaborar en un futuro, y que esta vez sea la buena. Clara, qué te voy a decir que no te haya dicho ya. Me abriste las puertas del despacho y me acogiste como uno más. Tú fuiste mi gran apoyo durante los primeros años. Me lo pusiste fácil y eso dice mucho. Aniol, la persona más constante que he conocido. Compañero, te auguro un futuro brillante allá donde vayas. Lisa, my favourite Dutch. Thinking of you is like thinking of sanity and a sense of humour. I'll remember you every Christmas while listening to The Killers carols. By the way... See? This weirdo is about to hand his thesis. Yay! Jaume, tú más que un compañero de trabajo eres un amigo. Gracias por aparecer en el momento justo. Gracias por escucharme hoy y todos los días. Un día me dijiste que había pasado a la categoría de las personas que te importan. Eso me dejó marcado. Estoy orgulloso de ti, padawan. Espera, ¿el padawan es el maestro o el aprendiz? A todos vosotros, las puertas de mi casa siempre estarán abiertas.

Me gustaría también agradecer a todos los miembros del VISCA y colegas su apoyo: Manel, eres toda una referencia en lo académico y en lo personal. Cada experimento que me explicas, cada cacharro que creas, cada charla aparentemente superficial... Ay, Manel. Qué voy a hacer ahora yo sin ti. ¡Si ya sabes que soy un desastre! Te haré caso, intentaré no volverme *demasiado* serio. Si algo he aprendido de ti es a disfrutar de lo que hago. A Jose Antonio por ser una especie de guía en la ciencia y apoyarme con las clases desde el primer día como profesor y como alumno. Antes de llegar a Barcelona ya te conocía de alguna manera. Creo que hablo en nombre de muchos cuando digo que tus apuntes de Percepción y Atención han ayudado a más de un estudiante de grado en otras universidades. A Matthias por enseñarnos las maravillas que se pueden hacer modelando una neurona. A Elisabet por acogerme nada más llegar a su lado y hacerme sentir como uno

más. A Àngels por estar siempre ahí para cualquier cosa que hiciera falta, incluso si lo que nos hacía falta era tener un mono con unos platillos en el despacho. A Javier por dar un toque de cordura, sensibilidad y memes a sobremesa. A Dani, por tu cercanía y ayuda desde el primer día. Me alegro de que por fin hayas vuelto. Y, por supuesto, a todos los demás miembros del departamento sobre los que podría extenderme páginas. Si hay algo que echo de menos en este momento tan raro son aquellas añoradas sobremesas.

Me gustaría hacer esta sección de agradecimientos extensible a toda mi vida universitaria, ya que creo que no estaría en este punto si no fuera por tres personas: Pilar, gracias por aquellas inspiradoras clases de Percepción y Atención. Sin duda durante aquellas clases decidí que mi futuro tenía que estar de alguna manera ligado a la investigación y a la percepción visual. Tu entusiasmo y tu cercanía fueron la chispa que me hizo interesarme por este tema. Cristina Muñoz, gracias por ser la persona más inteligente, sensible y curiosa que he conocido en mi vida. Eres la persona que más me ha hecho crecer personal y emocionalmente. Mi Sinsombrero favorita. Hemos pasado tantas cosas juntos que no puedo esperar a ver qué será lo siguiente. Por lo pronto, espero un largo paseo charlando por nuestro Madrid. Eli Brenner, you welcomed me and Cristina (both of them) at the VU with a big smile. There you showed us how far one can go with simplicity and excellent work. You are simply a role for me.

A mis padres, a quienes sin duda tengo que agradecer todo lo que soy. Me habéis enseñado desde pequeño a trabajar duro. A no conformarme. A ayudar a los demás y a intentar ser una buena persona cada día. Justo hoy papá se sorprendía cuando le daba las gracias por decirme que me apoyaríais fuera cual fuera la decisión que tomara sobre mi futuro. Eso es de agradecer, y mucho. No sabéis lo orgulloso que estoy de ser vuestro hijo. A Aldo y Alba a quienes echo de menos muchas más veces de las que se lo digo. Gracias por apoyarme desde la distancia.

A Àngel. Él me hace café por las mañanas, me ayuda con la resaca y me trae la bolsa de agua caliente mientras trabajo en esta tesis.

A amigos que han sido fundamentales durante estos años: Antonio, gracias por ser mi ancla y por ponerme los pies en el suelo tantas veces. Enrique, por enseñarme que nunca hay que dejar de aprender. A Sandra, por enseñarme qué es el esfuerzo y la dedicación por los demás. A Carlos,

por nuestras charlas interminables sobre cualquier tema random y por ser una continua fuente de inspiración intelectual. A Álvaro, Sandra, Alba y Juane por nuestros juegos de mesa. Recordad que lo importante en la vida es ponerse crema solar, comer pizzitas ricas e invertir en criptomonedas. A la próxima nos vemos en Karachi con Preciado.

A todos mis amigos: de Sonseca, de Madrid, a los que conocí por el mundo durante el Erasmus, a todos mis amigos del Máster, a los compañeros que están diariamente en la facultad, a las Flors, a los Corsegatos, a Xavi y Alba. Gracias por escucharme, por apoyarme y por darme miles de momentos inolvidables. Pero, sobre todo, gracias por no tener en cuenta las faltas y entender que no siempre he podido estar ahí física y mentalmente. Intentaré hacerlo mejor.

El desarrollo de esta tesis fue sustentado por la ayuda a la Formación del Profesorado Universitario (FPU17/01248) del Ministerio de Educación y Formación Profesional del Gobierno de España.

Summary

Until the present moment, an extensive amount of research has been done on how humans estimate motion or parameters of a task, such as the time-to-contact in simple scenarios. However, most avoid questioning how we extract 3D information from 2D optic information. A Bayesian approach based on a combination of optic and prior knowledge about statistical regularities of the environment would allow solving the ambiguity when translating 2D into 3D estimates. The present dissertation aims to analyse if the estimation of motion and time-to-contact in complex 3D environments is compatible with a combination of visual and prior information.

In the first study, we analyse the predictions of a Bayesian model with a preference for slow speeds to estimate the direction of an object. The information available to judge movement in depth is much less precise than information about the lateral movement. Thus, combining both sources of information with a prior with preference for low speeds, estimates of motion in depth will be proportionally more attracted to low speeds than estimates of lateral motion. Thus, the perceived direction would depend on stimulus speed when estimating the ball's direction. Our experimental results showed that the bias in perceived direction increased at higher speeds, which would be congruent with increasingly less precise motion estimates (consistent with Weber's law).

In the second study, we analyse the existing evidence on using a priori knowledge of the Earth's gravitational acceleration and the size of objects to estimate the time to contact in parabolic trajectories. We analysed the existing evidence for using knowledge of the Earth's gravity and the size of an object in the interaction with the surrounding environment. Next, we simulate predictions of the GS model. This model allows predicting the time to contact based on a combination of a priori variables (gravity and

ball size) and optic variables. We compare the accuracy of the predictions of time-to-contact with an alternative only using optic variables, showing that relying on priors of gravitation and ball size solves the ambiguity in the estimation of the time-to-contact. Finally, we offer scenarios where the GS model would lead to predictions with systematic errors, which we will test in the following studies.

In the third study, we created trajectories for which the GS model gives accurate predictions of the time to contact at different flight times but provides different systematic errors at any other time. We hypothesized that if the ball's visibility is restricted to a short time window, the participants would prefer to see the ball during the time windows in which the model predictions are accurate. Our results showed that observers preferred to use a relatively constant ball viewing time. However, we showed evidence that the direction of the errors made by the participants for the different trajectories tested corresponded to the direction predicted by the GS model.

In the fourth and final study, we investigated the role of a priori knowledge of the Earth's gravitational acceleration and ball size in estimating the time of flight and the direction of motion of an observer towards the interception point. We introduced our participants in an environment where both gravitational acceleration and ball size was randomized trial-to-trial. The observers' task was to move towards the interception point and predict the remaining flight time after a short occlusion. Our results provide evidence for using prior knowledge of gravity and ball size to estimate the time-to-contact. We also find evidence that gravitational acceleration may play a role in guiding locomotion towards the interception point.

In summary, in this thesis, we contribute to answering a fundamental question in Perception: how we interpret information to act in the world. To do so, we show evidence that humans apply their knowledge about regularities in the environment in the form of a priori knowledge of the Earth's gravitational acceleration, the size of the ball, or that objects stand still in the world when interpreting visual information.

Resumen

Hasta el momento, se ha realizado una gran cantidad de investigación sobre cómo el ser humano estima el movimiento o los parámetros de una tarea como el tiempo de contacto en escenarios simples. Sin embargo, la mayoría evita preguntarse cómo se extrae la información 3D a partir de la información óptica 2D. Un enfoque bayesiano basado en una combinación de información óptica y a priori sobre regularidades estadísticas del entorno interiorizadas en forma de conocimiento permitiría resolver la ambigüedad a la hora de traducir claves ópticas en 2D a estimaciones sobre propiedades del mundo en 3D. El objetivo de esta tesis es analizar si la estimación del movimiento y del tiempo de contacto en entornos 3D complejos es compatible con una combinación de información visual y a priori.

En el primer estudio, se analizan las predicciones de un modelo bayesiano con preferencia por las velocidades lentas para la estimación de la dirección de un objeto. La información disponible para juzgar el movimiento en profundidad es mucho menos precisa que la información sobre el movimiento lateral. Así, cuando se combinan ambas fuentes de información con un prior con preferencia por la velocidad baja, las estimaciones del movimiento en profundidad serán proporcionalmente más atraídas por el prior que las estimaciones del movimiento lateral. Por lo tanto, la dirección percibida dependería de la velocidad del estímulo. Nuestros resultados experimentales mostraron que el sesgo en la dirección percibida aumentaba a velocidades más altas, lo que sería congruente con estimaciones de movimiento cada vez menos precisas (consistente con la ley de Weber).

En el segundo estudio, analizamos las evidencias existentes sobre el uso del conocimiento a priori de la aceleración gravitatoria de la Tierra y el tamaño de los objetos para estimar el tiempo de contacto en trayectorias parabólicas. Analizamos las pruebas existentes sobre el uso del

conocimiento de la gravedad de la Tierra y el tamaño de un objeto en la interacción con el entorno. A continuación, simulamos las predicciones del modelo GS, un modelo que permite predecir el tiempo de contacto a partir de una combinación de variables a priori (gravedad y tamaño de pelota) y variables ópticas. Comparamos la precisión de las predicciones del tiempo de contacto con una alternativa que sólo utiliza variables ópticas, mostrando que basarse en las variables a priori de la gravedad y el tamaño de la bola resuelve la ambigüedad en la estimación del tiempo de contacto. Por último, mostramos varios escenarios en los que el modelo GS conduciría a predicciones con errores sistemáticos; escenarios que pondremos a prueba en los siguientes estudios.

En el tercer estudio, creamos trayectorias para las que el modelo GS da predicciones precisas del tiempo hasta el contacto en diferentes tiempos de vuelo, pero proporciona diferentes errores sistemáticos en cualquier otro momento. Hipotetizamos que, si la visibilidad de la pelota está restringida a una ventana de tiempo corta, los participantes preferirían ver la pelota durante las ventanas de tiempo en las que las predicciones del modelo son precisas. Nuestros resultados mostraron que los observadores preferían utilizar un tiempo de visualización de la pelota relativamente constante. Por otra parte, mostramos pruebas de que la dirección de los errores cometidos por los participantes para las diferentes trayectorias probadas se correspondía con la dirección predicha por el modelo GS.

En el cuarto y último estudio, investigamos el papel del conocimiento a priori de la aceleración gravitatoria de la Tierra y del tamaño de la pelota en la estimación del tiempo de vuelo y la dirección de movimiento de un observador hacia el punto de interceptación. Introdujimos a nuestros participantes en un entorno en el que tanto la aceleración gravitatoria como el tamaño de la pelota se asignaban aleatoriamente ensayo a ensayo. La tarea de los observadores consistía en desplazarse hacia el punto de interceptación y predecir el tiempo de vuelo restante tras una breve oclusión. Nuestros resultados proporcionan pruebas del uso del conocimiento previo de la gravedad y el tamaño de la pelota para estimar el tiempo de contacto. También encontramos pruebas de que la aceleración gravitatoria puede desempeñar un papel en la orientación de la locomoción hacia el punto de intercepción.

En resumen, en esta tesis contribuimos a responder a una cuestión fundamental en la Percepción: cómo interpretamos la información para actuar en el mundo. Para ello, mostramos evidencias de que los humanos aplican sus conocimientos sobre regularidades del entorno en forma de conocimiento a priori de la aceleración gravitatoria de la tierra, del tamaño de la pelota o de la estabilidad del mundo a nuestro alrededor para interpretar la información visual.

1 General Introduction

Contrary to what one might think, the world we live in is relatively predictable. As of the time of writing, it is clear that neither the keyboard will spontaneously grow like Alice in Wonderland nor start floating around the room in weightlessness. Although pushed towards the absurd, these examples bring out an important point. The world maintains an internal coherence that underpins predictability. Throughout this thesis, we will study how known statistics of the environment in the form of prior knowledge affect our percepts and help build predictive components that can be used to control our behaviour. Specifically, we will analyse how prior knowledge of relevant visual features affects motion estimation in different 3D environments leading to systematic errors in our perception.

This introductory chapter aims to give the reader the necessary context to understand the different experimental studies and grasp their relevance for motion perception and optic flow, that is, the retinal changing pattern. Thus, we must start with some key points. Throughout this thesis, we will assume a constructivist perspective of Perception. This approach assumes that recreate internal representations of features in the world based on sensory information (Helmholtz, 1867). These representations would allow us to predict future events based on simple computations that attempt to replicate common patterns in the real world. In our everyday life, these computations will allow us to predict that a cup will fall off the table if we bump it unintentionally. However, this prediction would not apply in a non-gravitational space, as the cup will be suspended in the air floating in weightlessness. As a result, the over-generalisation of these computational assumptions may lead to systematic errors.

From a probabilistic perspective, a feature about the visual world can

be represented in the form of a probability distribution. A *probability distribution* is a mathematical function that describes the occurrence of an event with two parameters: accuracy and precision. Accuracy would refer to the mean error of the estimation, while precision would refer to the variability of those errors. The following figure illustrates the combination of high/low accuracy and precision with an example of distance estimation (see Figure 1.1).

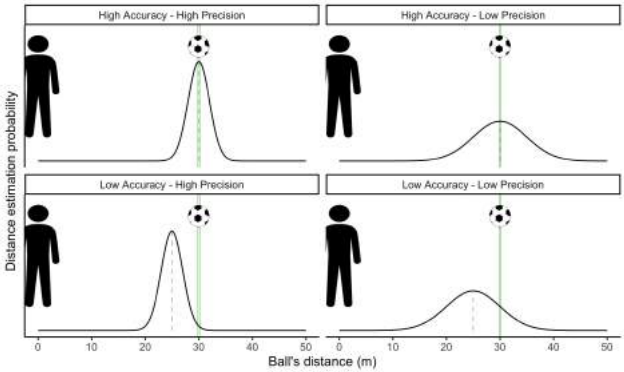

Figure 1.1. Representation of different levels of accuracy and precision (high/low). In this example, the ball is 30 meters from the observer (ball's position in x), denoted by a green vertical line. The y-axis represents distance estimation probability. The steeper the distribution, the more likely a ball distance will be estimated. Accuracy would represent the mean error concerning the real value to be represented. Precision would represent the variability of the errors.

In Figure 1.1, accuracy would refer to the average error between the actual distance to the ball (30 meters) and the estimates. If the estimates are systematically biased away from the actual distance (bottom panels), the accuracy of our estimate is low. Precision, however, would refer to the inverse of the variability in our estimates. Right panels represent a high variability in our estimates, that is, low precision. These two aspects become key when describing the prior distributions that reflect our perceptual knowledge of visual features of the environment. In turn, prior distributions affect how we solve a fundamental problem in perception: the inverse problem.

1.1 The inverse problem in Perception

Obtaining sensory information is key to grasping what is going on in our environment. However, the interpretation of sensory information has to serve some purpose. Whether it is to identify the nearest river or avoid a predator pouncing on us, being able to interpret our environment correctly, predict what will happen and act accordingly is key to maximising our survival. Here, the visual system is crucial, providing information about distant elements to interact with in the near future.

The problem with prediction is that we must first infer different states of the environment and then extrapolate them over time. As we will see below, the features of the objects around us are underspecified in the visual information reaching our retina. Hence, visual information does not suffice to infer those features in many cases. However, our ability to perform actions that require some degree of prediction, such as catching or hitting a ball on the fly, is astonishing (Brenner & Smeets, 2015; Regan, 2012). Therefore, understanding how we translate optical information into features of the world is fundamental.

As previously mentioned, relying only on visual information stimulating our retina presents a profound problem: using 2D images to infer properties of a 3D world. Relying on 2D images is problematic because there are many interpretations but just one real cause or distal stimulus. This epistemological problem is also known as the *inverse problem* of vision (Kersten et al., 2004; Pizlo, 2001). Figure 1.2 represents a simple example where many possible combinations of ball size and distance would project the same retinal angle onto the retina. The retinal angle is the projected angular size of an object on the retina (usually denoted by the Greek letter, θ).

It would be impossible to tell apart the actual size from the other possibilities without further information. Nevertheless, a simple solution to this problem can be found in many species: binocular vision. A binocular system provides two independent and horizontally displaced images. Their comparison solves the ambiguity partially within 2D images (Foley, 1980; Qian & Yazdanbakhsh, 2015). However, some elite sports players maintained their performance after losing sight with one eye (Regan, 2012).

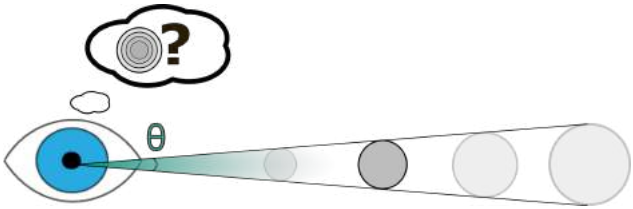

Figure 1.2. Representation of the inverse problem as a many-to-one problem, where many possible combinations of ball size and distances project the same angular size onto the retina. In this case, the observer has no means to infer the real combination of ball size and distance.

Thus, there might be another possible solution available with monocular vision. In the following equation, the reader can see how retinal size, object size and distance are related:

$$\text{Distance} \approx \frac{\text{Size}}{\theta} \tag{1.1}$$

Derived from the previous expression, one solution to this problem could be imposing constraints to the possible ball sizes or distances (Pizlo, 2001). This would be close to using a ball size as an *a priori*. If ball size is unknown or non-constant, distance cannot be computed. Thus, all distances would be equally possible. In Figure 1.3 an equiprobable estimation is represented by a uniform distribution (gold flat line). However, if ball size is known and constant, distance can be computed as the ratio between ball size and retinal size (Ittelson, 1951; Maltz et al., 2021). In that case, only a handful of distances would remain possible given our available information (black distribution in Figure 1.3). This simple example shows how prior knowledge of a contextual variable (ball size) and the assumption of a constant world, in this case, size constancy, provide the grounds to solve the inverse problem. However, this solution may not lead to a correct estimate if the wrong ball size is assumed.

Imagine that a Tennis ball size is assumed instead of assuming the correct size (Soccer ball). Following Equation 1.1 the distance would be underestimated (see Figure 1.3B). Thus, the accuracy of our prediction would be undermined. On the other hand, uncertainty in our estimation of distance would be caused by variability or noise in our visual measure,

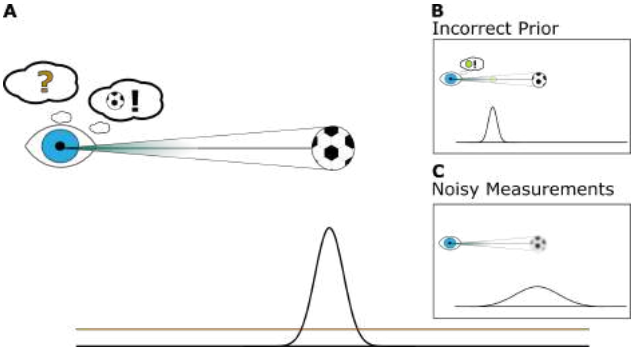

Figure 1.3. A) Probability distributions for judged distance with unknown ball size (gold line) and the correct ball size assumed (black line). B) Probability distribution assuming an incorrect ball size. C) Probability distribution with low precision visual measurements.

retinal size (see black distribution in Figure 1.3A). Perceptual uncertainty results from sensory measurements being affected by random noise of errors. Generally, the sensory representation of a physical quantity is finite given the limits of the biological structures involved in the process of transduction and communication processes within the nervous system (Faisal & Wolpert, 2009). In this sense, measurement variability will influence perceptual variability. For example, if the image of the ball is blurred, the retinal size would be more variable, which would translate into more uncertainty in distance estimation (see Figure 1.3C).

Here, classic psychophysics provides a standard method for estimating the discriminability (variability) of sensory measurements in our system (Fechner, 1948/1860). Weber's law indicates that the discriminability for most sensory modalities is a linear function of a constant known as the Weber fraction (usually referred to with the letter k). Comparing the Weber fraction across sensory modalities makes it possible to compare the precision with which they are represented. For example, the precision of positional estimates (Westheimer & McKee, 1977) is much higher than that of motion (de Bruyn & Orban, 1988; McKee, 1981) or acceleration (Gottsdanker et al., 1961a). As a rule of thumb, it is usually assumed that estimating variables of a higher temporal order (velocity compared with position) leads to increased measurement errors or Weber fractions.

This suggests that using variables such as acceleration within our estimates may not be the best approach because the following estimates will be very unreliable.

1.2 Perception as a Bayesian inferential process

Since Helmholtz (1867), Perception has been proposed as an inferential process in which visual information is interpreted as a best guess. This inferential process has been formulated in terms of "encoding" and "decoding" (Friston, 2010; Knill & Pouget, 2004; Wei & Stocker, 2015). *Encoding* would correspond to the activity resulting from the transduction of external energy that stimulates our sensory receptors resulting in an activity pattern in our nervous system. The information resulting from the encoding would be combined with prior knowledge during the *Decoding* providing a "read-out". The read-out represents an interpretation of the available data. In sum, the "encoding-decoding" framework generally assumes that we usually have prior knowledge that will assist during the decoding to solve the inverse problem of vision. In the following figure, the reader can see a scheme of the whole cycle involved in any sensorimotor task.

Combining new sensory evidence with our previous knowledge makes what is considered a Bayesian perspective. Following Helmholtz, 1867 thesis, the Bayesian perspective envisions Perception as an inferential process that combines all the available information to obtain the most likely interpretation (Maloney & Zhang, 2010). Within a Bayesian framework, all the available pieces of information are weighted and then combined, minimising the variance of the final estimate. The following expression represents a simplified version of the Bayes' theorem (Doya et al., 2007; Knill & Pouget, 2004), the algorithm that describes how prior and new information is combined.

$$\mathrm{P}(\mathrm{Event}|\mathrm{Data}) \propto \mathrm{P}(\mathrm{Data}|\mathrm{Event}) \times \mathrm{P}(\mathrm{Event}) \tag{1.2}$$

How does the above expression reads? The left-hand side of the expression is the Posterior, that is, the conditional probability of an event given the new information. In our perceptual experience, it would represent our percept. The right-hand side represents the Likelihood, the probability of

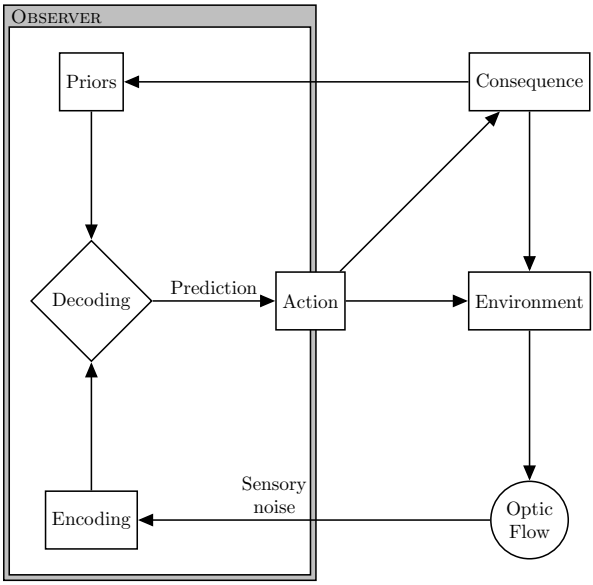

Figure 1.4. External visual information is *encoded* as a sensory array. Both sensory information and prior knowledge are combined within the *decoding* process. The product is a read-out or inference used to select an action from the existing repertoire. An observer's action would affect the environment and the optic array. From the difference between the expected outcome of the action and the consequences, the observer will store some information correcting the prior to refining further estimates.

the new information given an event multiplied by the Prior, that is, the probability of that event. Thus, the Posterior distribution would be proportional [1] to the product between the Likelihood and Prior distributions.

To describe the process of Bayesian inference, let us use an example. Imagine the context of a tennis training session. An athlete has to estimate the remaining flight time of a ball thrown by a well-known training machine. Her visual measurements (i.e. retinal size or the displacement of ball's image) would represent the Likelihood distribution for different flight durations Figure 1.5 (black lines in Figure 1.5). The previous experience with that machine would conform to the Prior knowledge (golden lines in Figure 1.5). The Likelihood and the Prior distributions will result in the Posterior. The Posterior distribution would represent the most probable flight durations given our available information (blue lines in Figure 1.5). The above example is an ordinary case where the player has access to reliable information (left panel in Figure 1.5). However, what would happen if visual information were compromised?

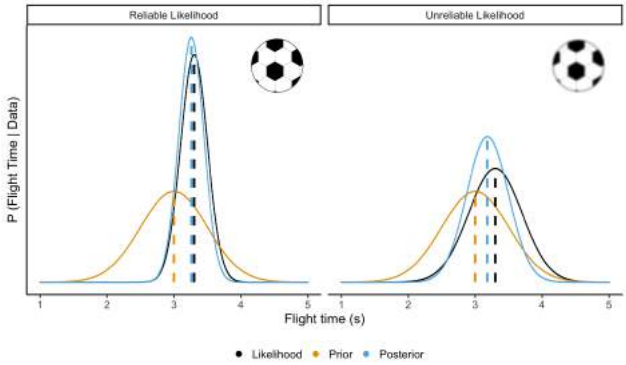

Figure 1.5. Representation of Bayesian estimation under the presence of reliable (left) and unreliable (right) sensory evidence (Likelihood). When the Likelihood is reliable, the effect of the Prior is limited. Thus, the Posterior is very close to the estimates provided by the Likelihood. In contrast, in the right panel, the Likelihood is unreliable. As a result, the Prior is weighted more heavily. Despite the Prior being the same, the relative weight differs due to different Likelihood reliability.

Imagine that the tennis player removes her contacts while training. In

[1] "$\propto$" reads as proportional.

this case, her visual information may be blurry due to myopia, and thus, her visual measurements would be unreliable. To optimise her flight time estimates, she should rely more heavily on her prior knowledge. In this case, relying more on the Prior shifted the Posterior distribution closer to the peak of the Prior distribution. Note that in both conditions, the Prior distribution is the same. However, relying more on the Prior if the Likelihood is unreliable would lead to an optimal estimate.

The Bayesian perspective has been a prolific research program in the last decades (see Geisler, 2011; Wei and Stocker, 2017 for reviews), showing that both sensorimotor tasks and perceptual judgments are consistent with combining visual and prior information. Among others, we would find pointing tasks (Trommershauser et al., 2003), manual interception (McIntyre et al., 2001; Zago et al., 2004), visual tracking tasks (Jörges & López-Moliner, 2020b), and motion judgements (Stocker & Simoncelli, 2006; Weiss et al., 2002) and temporal estimation (Jazayeri & Shadlen, 2010). However, Bayesian models have not gone uncriticised (Bowers & Davis, 2012; Rahnev & Denison, 2018). One of the main criticisms raised against the Bayesian perspective argues that many proposed models are underspecified. In this line, Bowers and Davis (2012) argues that usually, the parameters of either the Likelihood or the prior are chosen *a posteriori* to fit the data. Therefore, those models are hardly falsifiable. Nevertheless, recent works aimed at constraining better the Bayesian model to address this problem (Wei & Stocker, 2015, 2017).

1.3 A complex but predictable visual world

Framing Perception as a Bayesian inferential process is helpful because our world is governed by a series of regularities that we can extract from interacting with our environment. Throughout our lives, we encounter certain regularities as individuals or as species, such as the light coming from above (Adams et al., 2004), bigger objects being heavier (Peters et al., 2018) or objects of constant size (Ittelson, 1951; López-Moliner & Keil, 2012) that provide the grounds to solve the inverse problem on a day-to-day basis. However, in some cases, using those constraints would lead to perceptual errors or illusory percepts (Brunswik, 1956; Körding et al.,

2007; Samad et al., 2015). In the following sections, we will discuss how prior knowledge may assist in estimating motion and temporal predictions, highlighting how their use sometimes leads to perceptual errors.

1.3.1 Motion estimation

The study of local motion estimation has shown that perceived motion depends on factors such as motion direction, stimulus contrast, luminance or spatial frequency. For example, several studies have shown that motion for low contrast stimuli is consistently underestimated when compared with high contrast ones (Thompson, 1982; Weiss et al., 2002). A similar underestimation has been found for motion-in-depth compared to lateral motion (Brooks & Stone, 2006; Welchman et al., 2008). This bias may seem shocking at first glance. However, assuming that the objects around us generally remain still, a preference for slow motions arises as an optimal percept. This bias has proved to have general applicability, showing that it could even explain biases in the perceived direction of trajectories in a collision course (Aguado & López-Moliner, 2019; Welchman et al., 2008; Welchman et al., 2004), time perception (Chen et al., 2016) and even tactile illusions (Goldreich, 2007).

However, we generally do not estimate the speed of an object in the void. Instead, we estimate it in natural contexts for intercept or to predict the time it will take to reach some point. In most cases, this cannot be done directly using only visual information. Instead, it would be necessary to translate retinal measurements into 3D information, i.e. real-world measurements.

When we judge the velocity of two identical known balls at different distances moving at the same speed, we tend to perceive their movement as being equal even though the angular information is different (Distler et al., 2000; Rock et al., 1968; Zohary & Sittig, 1993) (see Figure 1.6A). This phenomenon is often referred to as speed constancy. Just as known object size allows to compute the distance with an object, it might also provide the grounds to estimate 3D motion in absolute coordinates. Indeed, speed constancy only holds in rich environments where contextual and pictorial cues are available. Instead, in a context with low pictorial cues such as a clear sky, speed estimates would be biased towards angular measures (Bian et al., 2013; McKee & Welch, 1989; Rushton & Duke, 2009).

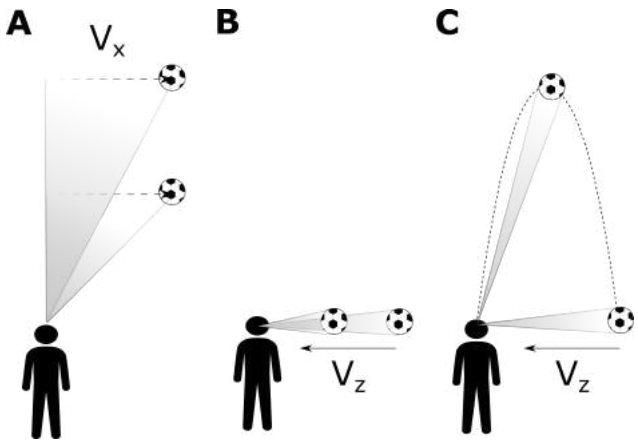

Figure 1.6. A) Two identical balls move at the same speed but at different distances, projecting distinct angular displacement. B) Ball moving in depth at a constant speed towards an observer. Motion-in-depth can be computed from retinal size correlates. C) Ball moving in a parabolic trajectory. Since distance does not change linearly, it can not be estimated from retinal size.

When an object moves towards the observer at a constant speed, motion in depth can be estimated using information about how the distance of the ball changes over time using retinal size correlates (see Figure 1.6B) (López-Moliner et al., 2007; Regan & Beverley, 1979; Rushton, 2004). However, in parabolic trajectories, the distance with the observer does not decline linearly. For example, in some cases, the distance to the ball increases as it travels upwards (see Figure 1.6C). For this reason, motion in depth cannot be retrieved through correlates of retinal size or binocular cues. Instead, some studies argue that a combination of known gravity and visual cues can provide estimates of motion-in-depth (Brouwer et al., 2006; Jörges & López-Moliner, 2017; Saxberg, 1987a). Nevertheless, the literature suggests that if possible, the observers would prefer to rely on correlates of retinal expansion and ball size to estimate motion-in-depth (Todd, 1981).

1.3.2 Temporal prediction

The time it takes an object to reach a predefined location is commonly known as the time-to-contact (TTC). Daily, we face scenarios where predicting precisely the time-to-contact with a car on the highway, a flying ball, or a cup falling off the table is vital to execute an action properly. Generally, position and velocity estimates can be used to extrapolate the movement. However, relying only on velocity estimates would result in significant temporal errors inconsistent with our own experience when dealing with accelerated motion.

One solution to intercept an object without prediction entails coupling some visual cue between actions and the motion of the object to be intercepted (Chapman, 1968; Montagne et al., 1999; Peper et al., 1994). This way, our actions would be self-corrected based on concurrent sensory information. However, relying on concurrent information may be problematic. Sensory feedback is usually delayed around 80-150 ms (Miall et al., 1986; Nijhawan, 1994). Because of that, online corrections would lead to jerky behaviour (Kistemaker et al., 2009; Tresilian, 1995) which is incoherent with amateur performance in batting or catching (McLeod et al., 2006; Regan, 1992).

Another solution that would avoid the need to predict could entail establishing mappings between different sources of visual information and temporal estimates. This strategy would allow producing timed actions without invoking internal computations or prior knowledge. For example, pressing a button when an object reaches a distance or an optic cue reaches a threshold (López-Moliner & Keil, 2012). This scheme would apply to a multitude of tasks even without sensorimotor feedback (Baurès et al., 2007; Zhao & Warren, 2015). Nevertheless, it would be a strategy attached to the conditions present when the mapping was learnt and thus, neither generalisable to other situations nor transferable to other tasks.

Thus, the prediction seems the most reasonable solution to reach such levels of accuracy at the time the system deals with sensorimotor delays. For example, we have previously mentioned that a preference for slow speeds could affect the estimation of the direction of an object in depth. This perceptual bias may impair our performance in catching or hitting a ball. However, continuously correcting our actions based on the perceived

errors as we obtain sensory feedback would explain our accuracy (Brenner & Smeets, 2018; Regan, 2012). Furthermore, previous studies have shown that the integration of new visual evidence may help us refine our motion (Snowden & Braddick, 1991) and temporal estimates (de la Malla & López-Moliner, 2015) in a Bayesian fashion similar to signal processing algorithms used in Engineering (Kalman, 1960).

However, predictions based solely on sensory information would not suffice to explain certain events in everyday life. For example, even if we only see the upwards part of a parabolic trajectory, we will predict the descending part (Grealy et al., 2004; López-Moliner et al., 2010). Incorporating visual acceleration estimates into the system would explain this phenomenon. However, our perception of acceleration is too poor to be used reliably (Gottsdanker et al., 1961a; Werkhoven et al., 1992). Thus, a strong prediction of the ball moving downwards can not be based on acceleration estimates. In turn, knowledge of the gravitational constant would be a way of incorporating acceleration into our predictions without the need to estimate it visually (Jörges & López-Moliner, 2017).

Research studying Gravity perception is challenging because visual and vestibular information may be inconsistent. For example, in Virtual Reality (VR), it is possible to modify the visual feedback to match arbitrary gravitational accelerations. However, the vestibular information from the otoliths would still be consistent with Earth's gravitational constant. A prominent exception is McIntyre et al. (2001). McIntyre et al. (2001) carried out the first experiment with astronauts in space where visual and vestibular information is congruent. Astronauts were able to time their actions accurately. However, they initiated the interceptive action consistently with an assumption of 1g. This effect persisted for the first few days of the experiment in the space, although the errors were slowly reduced over the days. These results show human attunement to Earth's gravity, which could be interpreted as a strong prior that adapts slowly to new environments.

Further studies have shown that the internal representation of terrestrial gravitation is consistent with tasks such as biological motion perception (Jokisch & Troje, 2003), interception of free-fall (Lacquaniti & Maioli, 1989) and parabolic trajectories (de la Malla & López-Moliner, 2015; Russo

et al., 2017) or judging durations (Jörges et al., 2021; Jörges & López-Moliner, 2019). Estimating the time-to-contact on parabolic trajectories would allow to anticipate the interception point or plan a displacement if needed. In this respect, the GS model (Gómez & López-Moliner, 2013) proposes an algorithm to predict the remaining time-to-contact of a parabolic trajectory with a combination of visual information and prior knowledge of gravitation or object size. Besides the a priori variables, the formulation of the GS model (Equation 4.3) includes monocular cues such as retinal size (θ), the elevation angle (γ), that is, the angle between eyes' and ball's height; and its the first derivative with time ($\dot{\gamma}$). In Figure 1.7, the reader can see a representation of the above-mentioned visual cues available in a parabolic trajectory.

$$TTC \approxeq \frac{2}{\mathrm{G}} \cdot \frac{\mathrm{size}}{\theta} \cdot \frac{\dot{\gamma}}{\cos(\gamma)} \tag{1.3}$$

The observer is assumed to use known constants of gravitational acceleration and object size. Therefore, this model would predict systematic errors that would be testable. As illustrated in Figure 1.7B, the time-to-contact is overestimated when faced with gravities greater than terrestrial. Conversely, the time-to-contact is underestimated with a ball bigger than the assumed size. Note that the direction of the errors with unexpected gravitations or ball sizes arise because each term is in the denominator and the numerator of the expression, respectively.

This model predicts the time-to-contact accurately for purely parabolic trajectories in a collision course with the observer. Nevertheless, in real life, parabolic trajectories are deviated by air drag or the Magnus effect (Adair, 2002; Brancazio, 1985; Hubbard, 1995) which would undermine the accuracy of the predictions.

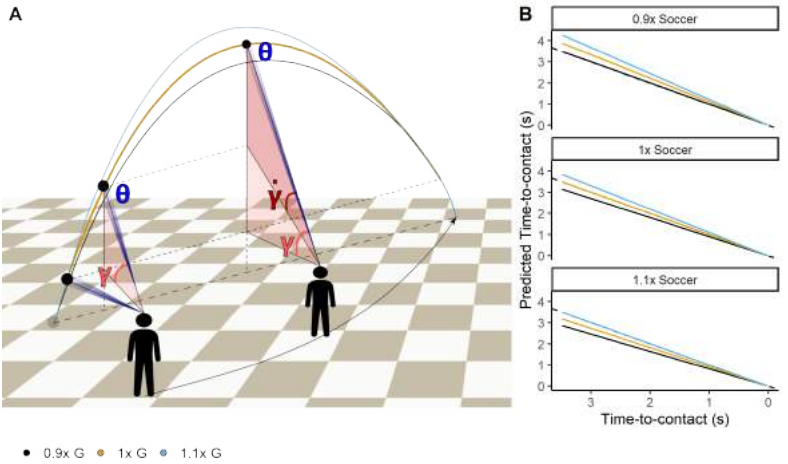

Figure 1.7. A) Representation of a parabolic trajectory under different gravitational accelerations (see legend). The monocular cues available for the estimation of time-to-contact are represented with shaded areas. Retinal size (θ, in blue), the elevation angle (γ, in pink) and the rate of change of the elevation angle ($\dot{\gamma}$, in burdeos). B) Predictions of time-to-contact for head-on approaches based on the GS model. A discontinuous line in the background indicates perfect accuracy in the predictions. Terrestrial gravity (G = 9.807 m/s) and Soccer ball size were assumed. Trajectories with different ball sizes (panels) and gravitational accelerations (colour code) were simulated.

2 Research objectives

2.1 Aims and objectives

During the introductory chapter, we have discussed one of the fundamental problems of Perception, the inverse problem and how the use of prior distributions can tackle this problem. Studying how information is processed to solve the inverse problem is the foundation behind the main objective of this thesis.

So far, extensive research has been carried out on estimating motion or time-to-contact in relatively simple scenarios. However, most of these studies have avoided questioning whether we can extract 3D information from visual information when faced with tasks like catching a ball on the fly. A Bayesian approach provides predictions based on visual information and prior knowledge. The main objective of this thesis is to analyse whether estimating the motion and time-to-contact of objects in complex 3D environments is compatible with a Bayesian model. In order to analyse our main objective, we set different sub-objectives:

- **Objective 1:** Analyse whether direction estimation is consistent with a Bayesian model with a preference for slow speeds and the estimates of 3D motion following Weber's law.

- **Objective 2:** Describe the available evidence on the use of priors about object gravitational acceleration and object's size. Simulate and analyse the predictions of time-to-contact of a model based on both priors (GS model) to characterise their accuracy and precision in complex environments.

- **Objective 3:** Analyse whether the errors in estimating time-to-

contact for parabolic trajectories are consistent with predictions based on the GS model.

2.2 Hypothesis

2.2.1 Motion direction estimation with motion-in-depth

As explained in the introductory chapter, the estimation of the direction of an object in collision with the observer is commonly biased. The existing explanation for this bias would be consistent with the combination of visual sensory information and a prior expectation of zero motion in the scene. Specifically, this bias would arise from the differential attraction towards the prior for lateral and depth movement. However, to further constrain and validate this model, it would be necessary for this perceptual bias to occur under the same prior but with different levels of physical and perceived speed. According to Weber's law, trajectories with higher velocities will experience reduced precision, which, in turn, would result in more significant biases in the perceived direction. In Chapter 3 (Study 1), we worked with two main hypotheses to test if the estimation of direction is consistent with a Bayesian model on motion estimation:

- **Hypothesis 1.1:** Trajectories of larger physical and perceived speed will lead to greater biases in perceived direction.
- **Hypothesis 1.2:** A slow-motion prior will explain direction estimates will less error than a model based solely on visual information (e.g. visual evidence).

2.2.2 Gravity and Size Priors to estimate the Time-to-contact

To address the second objective of this thesis, in Chapter 4 (Study 2), we review the literature supporting the use of internalised priors of size and gravity for the estimation of time-to-contact in 3D parabolic trajectories. We worked with the hypothesis that (**Hypothesis 2.1**) the predictions of time-to-contact for parabolic trajectories based on the GS model are more accurate and precise than the predictions based only on visual information.

We first reviewed the optic cues readily available for an observer from a parabolic trajectory to estimate the time-to-contact. Our literature review suggests that all the variables included in the computations of the GS model are encoded precisely enough to provide a reliable estimation of time-to-contact in combination with priors of gravity and ball size. Then, we simulated predictions based on the GS model and analysed their accuracy and precision compared to predictions based on a combination of optic cues. Our simulations sketch the GS model as a flexible algorithm that provides predictions of the remaining time-to-contact even when the ball's course does not follow a perfect parabola or the ball is not in a collision course with the observer.

2.2.3 Time-to-contact in 3D parabolic trajectories

In the course of Chapter 4 (Study 2), we showed that the GS model predicts systematic errors in different situations:

- Trajectories directed to a point other than the observer's location.
- Trajectories in which the assumed gravity or ball size do not match the actual parameters of the task.
- Trajectories affected by complex effects such as air drag.

In Chapter 5 (Study 3), we analysed the first situation where the trajectory is not in a collision course with the observer. In Chapter 4 (Study 2), we described how the predictions of the GS model would differ depending on the ending point of the trajectory with respect to the observer. Concretely, we identified that for trajectories directed to a point behind the observer, the estimates of the time-to-contact are accurate at two different moments during the flight. At launch and at a different time during the flight depending on the specific geometry of the parabola. Concretely, when the viewing angle describes a right angle between the observer, the ball and the interception location. If an observer were to use the GS model to predict time-to-contact but viewing time is restricted to a brief temporal window, she would benefit from using the information within the time window when the predictions are accurate. During any other viewing time, the predictions of the GS model would present systematic errors. Those errors would

allow us to compare with the errors in our participants' estimates. Taking into account the above, in this Chapter 5 (Study 3), we worked with two different hypotheses:

- **Hypothesis 3.1:** If participants exploit the GS model, they will look at the trajectory more often when the predictions of the GS model are more accurate.

- **Hypothesis 3.2:** Estimates of the remaining time-to-contact will be in the direction expected by the GS model across different trajectories.

In Chapter 6 (Study 4), we study the second situation mentioned above in which the GS model would predict systematic temporal errors: when the gravity and ball size do not correspond to the standard *expected* values. In this study, gravitation and ball size were arbitrarily selected and thus, would not correspond with our participants' expectations. We designed a task in which our participants had to move in a virtual reality setup replicating the catch of a flying ball and predict the time-to-contact after an occlusion at the end of the trajectory. We hypothesised that our observers would commit systematic errors in the direction predicted by the GS model under unexpected gravitations and ball sizes. Specifically, we hypothesise that:

- **Hypothesis 4.1:** Our observers would overestimate the remaining time-to-contact if gravitation is greater than the standard on Earth and vice-versa.

- **Hypothesis 4.2:** Our observers would underestimate the remaining time-to-contact if the ball size is larger than the standard of a Soccer ball and vice-versa.

- **Hypothesis 4.3:** Our observers would deviate their path travelled towards the interception location in a direction consistent with a misestimation of time-to-contact due to unexpected values of gravity and ball size.

3 Study 1: Perceived speed of motion in depth modulates misjudgements of approaching trajectories consistently with a slow prior

Previous studies have shown that the angle of approach is consistently overestimated for approaching (but passing-by) objects. An explanation based on a slow-motion prior has been proposed in the past to account for this bias. The mechanism relies on the (less reliable) in-depth component of the motion being more attracted towards the slow motion prior than the (more reliable) lateral component. This hypothesis predicts that faster speeds in depth will translate into a greater bias if the perception of velocity in depth follows Weber's law. Our approach is different than the one used in previous studies where perceived speed and direction were measured in different experiments. To test our hypothesis, we conducted an experiment in which participants estimated approaching angles via a pointing device, while at the same time comparing the speed of the approaching object with a lateral velocity reference. This way, we couple perceived speed with perceived trajectory for each approaching angle in the same trial. Our results show that the directional bias is larger for faster objects, which is consistent with motion in depth following Weber's law. The differential biases can be accounted for by a Bayesian model that includes a slow motion prior.

This study has been published as: Aguado, B., & López-Moliner, J. (2019). Perceived speed of motion in depth modulates misjudgements of approaching trajectories consistently with a slow prior. *Vision Research*, *159*, 1–9

3.1 Introduction

One of the main functions of the visual system is to recover the 3D structure of the environment. This is particularly important when we need to estimate the direction and speed of moving objects on a collision (or near-collision) course with us.

Knowing how different cues, both monocular and binocular, contribute to estimate direction and motion in depth (MID) has attracted interest in the past (Beverley & Regan, 1973; Cumming & Parker, 1994), but still is an active field of research (J. Harris et al., 2008; Rokers et al., 2018). Past work on MID, however, has mainly focused on precision and accuracy of motion estimates (J. Harris & Dean, 2003; Rushton & Duke, 2009).

Regarding the direction of approach, several studies have shown that we tend to overestimate the bearing angle (from now on β); see Figure 3.1) of the trajectory of an approaching target (J. Harris & Drga, 2005; Lages, 2006; Poljac et al., 2006; Welchman et al., 2004). This is, we overestimate the lateral distance by which a ball passes us. This can be counter-intuitive, given that we are very sensitive to the motion direction of objects on a collision course (Regan et al., 1986).

To explain this bias, Welchman et al. (2008) put forward a Bayesian explanation that included the so-called *Slow Motion Prior* which is a main component of a motion perception model by Stocker and Simoncelli (2006). Sensory estimates (likelihood) are combined with an expectation of nearly zero motion in the environment (prior) resulting in consistent underestimations of speed (posterior), with the extent of underestimation depending on the reliability of the likelihood (e.g. contrast of a grating; Stocker and Simoncelli, 2006). Therefore, if the reliability of the signal is low, the slow prior will be weighted more, resulting in a slower posterior and, consequently, the speed of the stimulus will be underestimated more strongly.

In the same study, Stocker and Simoncelli (2006) found that the width of likelihood estimates for speed discrimination tasks increases logarithmically as a function of speed approximately following Weber's law (for targets moving faster than 1 deg/sec), as suggested by previous literature in the field (McKee et al., 1986; Welch, 1989). Furthermore, they used

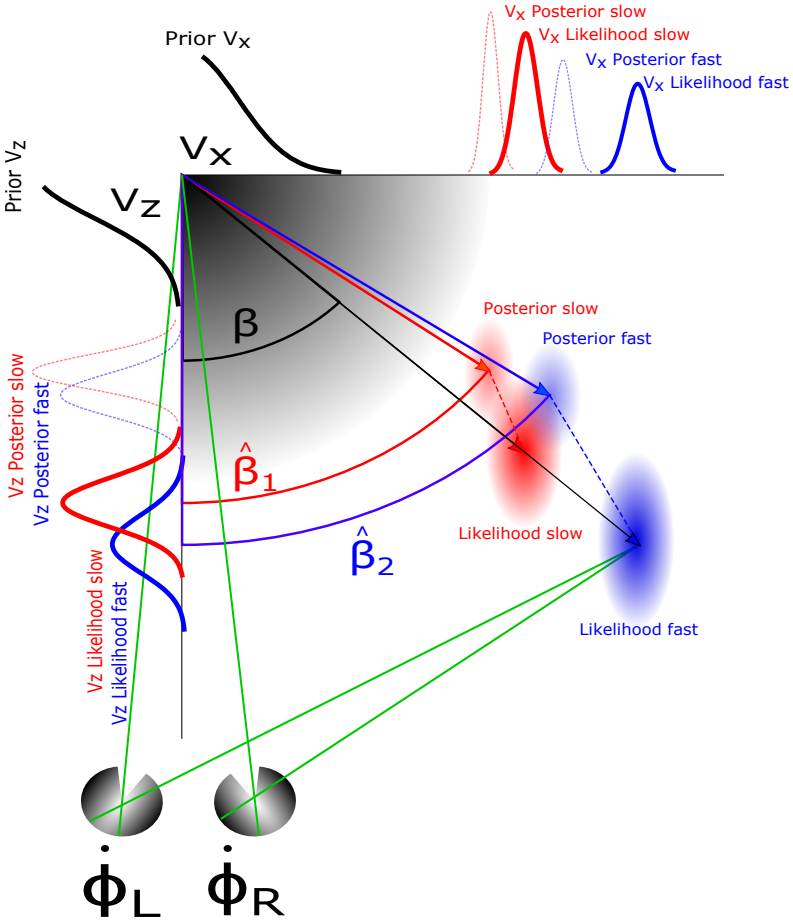

Figure 3.1. The prior is represented by the grey radial gradient centred at $V_x = V_z = 0$. Two different movements with the same bearing angle β are depicted in this scene. The speed of each movement is indicated by the color of the arrow: the slow movement is indicated in red whereas the fast movement is indicated in blue. Assuming Weber's law, the faster movement is noisier than the slower one, which is denoted by the SD of the respective likelihood gaussian ones. In addition, V_z is less reliable than V_x (depicted by the spread of the likelihood distributions for each vector, represented as thick lines at the margin). The effect of the prior (grey radial gradient, centered at $V_x = V_z = 0$) affects each speed vector differently. This effect is represented by the shift of the posterior distribution (distributions represented with dotted lines for each vector at the margin). Given that the prior will affect the slow and fast movements differently, the perceived trajectory would depend on the physical speed of the movement while keeping the physical trajectory constant (β). The perceived trajectory is denoted by $\hat{\beta}_1$ for a slow movement (red) and $\hat{\beta}_2$ for a fast movement (blue). The dashed segments connecting the centroids of likelihood and posteriors denote the speed bias for each movement. $\dot{\phi}$ represents rate of azimuth change for each eye.

a Bayesian Observer model to infer the shape (SD) of the Slow Motion Prior, which falls from a peak at slow speeds becoming shallower for faster ones. As a result, the prior expectation introduces increasingly biases for the posterior as a function of the perceived speed.

Welchman et al. (2008) explained the underestimation of approaching angles in terms of this slow prior: The estimate of the lateral component (V_x); Equation Equation 3.1) is more reliable than the estimate of the depth component in MID (V_x); Equation Equation 3.1)(see Figure 3.1).

$$V_x \approx \frac{\dot{\phi} \cdot \mathrm{i} \cdot \mathrm{d}}{\mathrm{i} + \dot{\delta} \cdot \mathrm{d}} \tag{3.1}$$

$$V_z \approx \frac{\dot{\delta} \cdot \mathrm{d}^2}{\mathrm{i} + \dot{\delta} \cdot \mathrm{d}} \tag{3.2}$$

V_x estimates are based on the rate of change of the azimuth ($\dot{\phi}$); Equation Equation 3.3; Figure 3.1) and the rate of change of the disparity ($\dot{\delta}$); Equation Equation 3.4), while V_z depends solely on the rate of change of the disparity. The variance of the azimuth signal is up to two times lower than for disparity. As a consequence, V_x, as it combines both signals, is much more reliable than V_z, which only depends on the less reliable signal, the rate of change of the disparity ($\dot{\delta}$)) (see Gaussian curves in Figure 3.1). In Equation 3.1 and Equation 3.2: i stands for interocular distance and d for viewing distance.

$$\dot{\phi} \approx \frac{\dot{\phi}_L + \dot{\phi}_R}{2} \tag{3.3}$$

$$\dot{\delta} \approx \dot{\phi}_L - \dot{\phi}_R \tag{3.4}$$

As Stocker and Simoncelli (2006) pointed out, the greater the reliability of the measure, the lower the effect of the prior and vice-versa (see Figure 3.1). As a consequence, the fact that V_x is much more reliable than V_z should result in a differential influence of the prior for the posterior speed estimate. Under the assumption that Weber's Law holds for motion in depth, yields the prediction that faster velocities should lead to more biased trajectories.

Specifically, our hypothesis suggests that higher speeds in depth would

lead to a larger overestimation of the bearing angle. Since we assume that the perception of the trajectory depends on the perceived speed (i.e. how the depth component is affected by the slow motion prior), we can test the hypothesis that the bias can be accounted for by how different speeds are encoded through the mediation of a slow prior. Thus, we investigated if different perceived speeds in depth lead to different degrees of bias on the perceived angle. To do so, we followed a different approach than Welchman et al. (2008). In the present study, we asked for the perceived speed in the same trials used for adjusting the direction. In this way, we can study if perceived speed is related to perceived direction for the same physical trajectories. In this study, only the case in which the initial lateral position of an object is the same as the position of the observer will be taken into account. For a more general approach, see (Rokers et al., 2018). As a final step, we formulated a Bayesian model to test the predictions of a slow motion prior model.

3.2 Methods

3.2.1 Participants

Eleven observers participated in the experiment. All of them had normal or corrected-to-normal vision and were naive with respect to the purpose of the experiment. One subject was stereo-blind as tested with StereoFly test (Stereo Optical Co.) and had to be excluded from further analysis. The final sample consisted thus of ten participants (n=10).

The research in this study is part of an ongoing research program that has been approved by the local ethics committee of the University of Barcelona. This study is in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Before taking part in the experiment, every subject signed an informed consent form.

3.2.2 Material

Apparatus

Two Sony laser projectors (VPL-FHZ57) were used to provide overlaid images on a back-projection screen (244 cm height and 184 cm width) with a resolution of 1920x1080 pixels. The refresh rate of the image was 85 Hz for each eye. Circular polarizing filters were used to provide stereoscopic images. A pointing device (see Procedure; Figure 3.2) was used to record the perceived direction with the position data being acquired by an Arduino board. The device was calibrated to provide linear measures of azimuth within a 180 degrees space. It was calibrated only once to make sure that the measurements were consistent across the participants. A custom input device with 2 buttons was used for the participants to indicate whether the first or second ball moved faster within each trial.

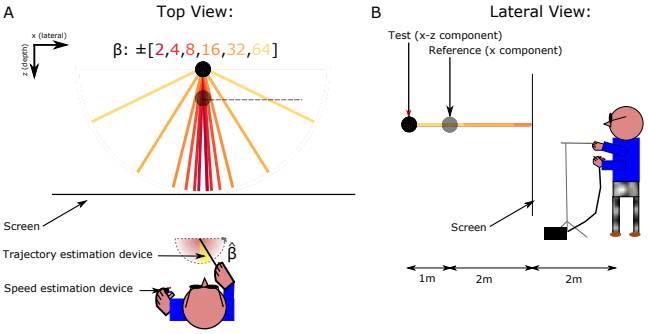

Figure 3.2. The figure represents a top (A) and lateral (B) view of the setup. (A) The colour of each trajectory indicates the bearing angle. In order to produce an estimate of the bearing angle, the participants aligned the pointer with the perceived direction of movement. To estimate the speed of the target moving in depth the participants were instructed to press a button indicating which ball had moved faster (blue for first/reference or red for second/test). As illustrated in the top view, the reference target only included a lateral component of movement (grey dashed line). However, test target included depth and lateral speed components (no vertical speed component was involved)

Stimulus

The stimulus consisted of two spheres with a checked texture presented consecutively. The first ball (reference target) moved along the x axis without any depth or vertical component. The second ball (test target) included a z (depth component) approaching to the observer at different bearing angles (β = 2, 4, 8, 16, 32, 64 degrees) which would pass the observer either to their left (negative β) or right (positive β) (see colour code for each β in Figure 3.2). The simulated vertical position of the target was 1.48 meters above the ground. The vertical speed component was 0, that is, there was no change in vertical position. The initial distances from the observer were 4 and 5 meters for the reference and test target respectively. We chose these values to match the average visual angle for both targets. The physical radius of the target was 3.3 cm (the size of a standard tennis ball). The presentation time was fixed to 1 s. The reference speed and β were pseudo-randomized within participants and angles taking the values 2, 2.5 or 3 m/s (26.5, 32 and 37.5 deg/s). Once determined, the reference target was the same for a given angle and participant, while the speed of the test target (with depth component) varied according to a Bayesian staircase (see Procedure). Since the estimated duration of the experiment using all possible combinations of direction and velocity (3 speeds and 12 directions) was too high, each participant observed only 12 different pseudo-randomized conditions.

3.2.3 Procedure

The experiment was performed in a dim room. Participants stood centrally at 2 m distance from the screen. The projected disparity was adapted to each participant's inter-ocular distance.

One session consisted of 360 trials (30 trials per approaching angle). Each trial consisted of two trajectories. We first showed the reference target that moved laterally (V_x component only: reference speed), followed by the test target (test speed: motion in depth with both V_x and V_z components). The speed of the test target varied according to a QUEST procedure (A. B. Watson & Pelli, 1983). We ran a total of twelve interleaved QUEST staircases (30 trials each), one for each combination of speed and approaching

angle to compute the point of subjective equality (PSE) of speed between the test and reference target, that is, the speed of the motion in depth that is perceived as fast as the reference movement.

The participants completed 10-15 training trials previous to the main experimental procedure in order to familiarize themselves with the task. No feedback about response performance was provided during any part of the experiment.

After the two stimuli disappeared, an auditory signal prompted the participants to perform 2 different tasks in each trial:

1. *Trajectory estimation*: The observers were required to estimate the trajectory of the test target with the pointing device. To this end, the participants aligned the pointer with the perceived direction of the movement.

2. *Speed estimation*: In a two alternative forced choice (2AFC) task, the participants were instructed to press a button indicating which ball had moved faster (left for first/*reference* or right for second/*test*).

The participants gave their response for the speed estimation task while keeping the pointing device aligned with the estimated trajectory (from now on $\hat{\beta}$), such that both responses were registered simultaneously.

3.3 Model Specification

We developed a Bayesian model to estimate the variability of a prior in the $x - z$ plane that can describe the perceived trajectories in our experimental results. To define this model we assumed that motion and direction were estimated consistently with one another. For the sake of simplicity we introduced the model assuming that $\beta > 0$, but the same would apply for $\beta < 0$. Stocker and Simoncelli (2005) found that motion estimates deviated from the Weber's law for low angular speeds (< 1 deg/s). In order to account for this deviation, they decided to introduce a correction for speed estimates in their model. Even though Welchman et al. (2008) introduced this adjustment, since the angular velocities of our experiment are much greater, we have not corrected for this deviation.

We defined the likelihood distribution of our test stimulus as a 2D Gaussian (N) for each β, that is, a join of two 1D distributions, one for each axis or component (i.e. V_x and V_z as viewed by an observer).

$$p(\dot{\phi}|V_x, V_z) = N_{(\mu_{xz}, \sigma^2_{xz})} \tag{3.5}$$

$$p(\dot{\phi}|V_x) = N_{(\mu_x, \sigma^2_x)} \tag{3.6}$$

$$p(\dot{\phi}|V_z) = N_{(\mu_z, \sigma^2_z)} \tag{3.7}$$

Means for V_x and V_z were defined from the estimated PSE for each reference speed and β:

$$\text{PSE}_x = \text{PSE} \cdot \ sin(\beta) \tag{3.8}$$

$$\text{PSE}_z = \text{PSE} \cdot \ cos(\beta) \tag{3.9}$$

Standard deviations for V_x and V_z were defined as the discrimination thresholds for each component. Assuming that Weber's law holds for the range of speeds shown in this experiment, we calculated the discrimination thresholds for each component. Lateral discrimination thresholds were calculated assuming a Weber fraction of 10% as previously found in the literature (Portfors-Yeomans & Regan, 1996; Welchman et al., 2008). Depth discrimination thresholds, similarly, were obtained assuming that the Weber fraction at $\beta = 2$ is approximately equivalent to that of a stimulus in a collision course with the observer ($\beta = 0$). Therefore, we obtained the standard deviations from the respective Weber fractions (from now on WF for x and z axis by β and participant (see inset at Figure 3.2B):

$$\sigma_x = WF_x \cdot V_{Ref} \cdot sin(\beta) \quad \Rightarrow \quad \sigma_x \approx 0.1 \cdot V_{Ref} \cdot sin(\beta) \tag{3.10}$$

$$\sigma_z = WF_z \cdot V_{Ref} \cdot cos(\beta) \quad \Rightarrow \quad \sigma_z \approx WF_{\beta \pm 2} \cdot V_{Ref} \cdot sin(\beta) \tag{3.11}$$

Therefore, V_x and V_z follow a 2D Gaussian distribution with components:

$$p(\dot{\phi}|V_x) = N_{(PSE_x,\sigma_x^2)} \tag{3.12}$$

$$p(\dot{\phi}|V_z) = N_{(PSE_z,\sigma_z^2)} \tag{3.13}$$

The distribution of the prior was defined as an isotropic 2D Gaussian in real world speed with mean 0 and a free parameter (variance). We chose a Gaussian with mean 0 since, in practical terms, a Gaussian approximates very well to a prior distribution following a power law as used by Stocker and Simoncelli (2006) and Welchman et al. (2008) ($p(|v|) = exp^{-7.04(\dot{x}^2+\dot{z}^2)-8.49}$). Additionally, this is also motivated by the assumption that speeds > 1 (deg/s) follow Weber's law.

$$p_{v_x} = N_{(0,\sigma_v^2)} \tag{3.14}$$

$$p_{v_z} = N_{(0,\sigma_v^2)} \tag{3.15}$$

The variance of the prior (${v_\sigma}^2$) is the only free parameter in this model. To estimate it, we obtained the posterior of each component $\hat{V}_x$ and $\hat{V}_z$ by means of a MLE procedure (Ernst & Banks, 2002). This procedure combines different sources of information (prior and likelihood) in a weighted fashion to obtain an optimal posterior estimate (see Figure 3.1):

$$\hat{V}_{x_{Pred}} = \frac{PSE_x}{1 + \left(\frac{\sigma_x}{\sigma_v}\right)^2} \tag{3.16}$$

$$\hat{V}_{z_{Pred}} = \frac{PSE_z}{1 + \left(\frac{\sigma_z}{\sigma_v}\right)^2} \tag{3.17}$$

By simple trigonometry, the trajectories were calculated as a function of the predicted components of speed.

$$\hat{\beta}_{Pred} = arctan\left(\frac{\hat{V}_{x_{Pred}}}{\hat{V}_{z_{Pred}}}\right) \tag{3.18}$$

As a result:

$$\text{Bias}_{\text{Pred}} = \hat{\beta}_{\text{Pred}} - \beta \tag{3.19}$$

We estimated the free parameter (prior variance) in a optimization routine using the *optim* function included in R (R Core Team, 2020). The objective function was formulated as a minimization of the sum of squared differences between the predicted ($\hat{\beta}_{\text{Pred}}$) and the measured trajectories ($\hat{\beta}$) for each trial.

$$Min : \sum (\hat{\beta}_{Pred} - \hat{\beta})^2 \tag{3.20}$$

Substituting in Eq. Equation 3.18 with Equation 3.16 and Equation 3.17:

$$\hat{\beta}_{Pred} = arctan\left(\frac{\text{PSE} \cdot sin(\beta) \,/\, (1 + (WF_x \cdot V_{Ref} \cdot sin(\beta)/\sigma_v)^2)}{PSE \cdot cos(\beta) \,/\, (1 + (WF_z \cdot V_{Ref} \cdot cos(\beta)/\sigma_v)^2)}\right) \tag{3.21}$$

In Figure 3.5B we use Equation 3.21 to show the predicted perceived trajectories for each reference speed using parameters $WF_x = 0.1$; $WF_z = 0.28$; and the prior standard deviation obtained by the optimization routine $\sigma_v = 0.33$.

3.4 Data analysis

3.4.1 Speed estimation

We fit a cumulative Gaussian curve (mean and SD) to the proportion of faster than standard responses for each β and participant using the R (R Core Team, 2020) package *quickpsy* (Linares & López-Moliner, 2016) in order to obtain the PSE (i.e speed in the cumulative Gaussian curve that elicited 50% *faster from the standard*). We defined the discrimination thresholds as the half difference between 16% and 84% *faster from the standard* response probabilities in the psychometric function (i.e. 1 standard deviation).

Then, we calculated the Weber fractions for each β and reference speed

as the discrimination threshold divided by the speed of the reference motion. We calculated this value for each participant and bearing angle separately.

Next, we obtained a ratio between the PSE and the reference speed. This value represents a measure of the degree of underestimation of the test relative to the reference speed. Ratios larger than 1 would denote an underestimation of test speed compared to reference speed.

Then, we used the PSE as a boundary to classify each trial as perceived as *slow* and *fast*. We made sure that the mean speed difference between both categories (i.e. *slow* and *fast*) was above discrimination threshold for speed by comparing the discrimination threshold with the speed difference between *slow* and *fast* groups across β in a two-way ANOVA.

3.4.2 Trajectory estimation

First, we filtered out those trials in which participants misjudged the absolute direction of the test movement (either left or right; less than 1%). Then we fitted a Linear Mixed Model to disentangle the effect of the speed group and β over $\hat{\beta}$ with the R-package *lme4* (Bates et al., 2015). We transformed the dependent variable $\hat{\beta}$ into its absolute value and β into the logarithm of its absolute variable to linearize the data. Speed group, β and their interaction were introduced as fixed effects. Slopes of β by participant and trial were introduced as random effects.

3.5 Results

3.5.1 Speed estimation

Figure 3.3A shows the psychometric fit for the speed judgement of a representative participant. In this example, both psychometric fits were carried out for the same reference speed (2 m/s), but for two different bearing angles (β). Figure 3.3B illustrates to what extent speed in depth was underestimated relative to lateral speed (ratios larger than one denote underestimation of the test target speed). Speed in depth was strongly underestimated for trajectories closer to the observer (smaller β). For these trajectories, the velocity had to be increased by a factor of 1.5-2 in order to be perceived as moving as fast as the reference. The underestimation

Table 3.1. Weber fractions for each β.

	Weber Fraction	CI
-64	0.13	0.06 - 0.20
-32	0.10	0.02 - 0.17
-16	0.14	0.02 - 0.26
-8	0.20	0.11 - 0.29
-4	0.26	0.14 - 0.38
-2	0.23	0.09 - 0.37
2	0.30	0.14 - 0.46
4	0.22	0.14 - 0.31
8	0.08	0.04 - 0.12
16	0.16	0.10 - 0.22
32	0.08	0.03 - 0.14
64	0.13	0.01 - 0.24

Table 3.2. Weber fractions for each reference speed.

	Weber Fraction	CI
2.00	0.18	0.14 - 0.23
2.50	0.16	0.12 - 0.21
3.00	0.16	0.12 - 0.20

attenuates and disappears for trajectories approaching a lateral movement. These results are in line with the matching speeds for lateral and depth motion in Welchman et al. (2008).

Figure 3.3C shows Weber fractions for the estimation of speed in depth. As expected, Weber fractions depend on the presented trajectory ($F(11, 85) = 2.73$, $p = .005$) and increased as the V_z component (i.e. β) increased. Participants thus judged the speed of MID less accurately than for lateral motion (see Table 3.1). Our results show a Weber fraction close to 0.25 for depth speed estimation, which is in agreement with those values reported by Welchman et al. (2008) and Rushton and Duke (2009). However, we found no significative effect of the reference speed ($F(1, 85) = 0.001$ $p = .975$, see Table 3.2) or interaction effect ($F(11, 85) = 0.91$, $p = .538$)

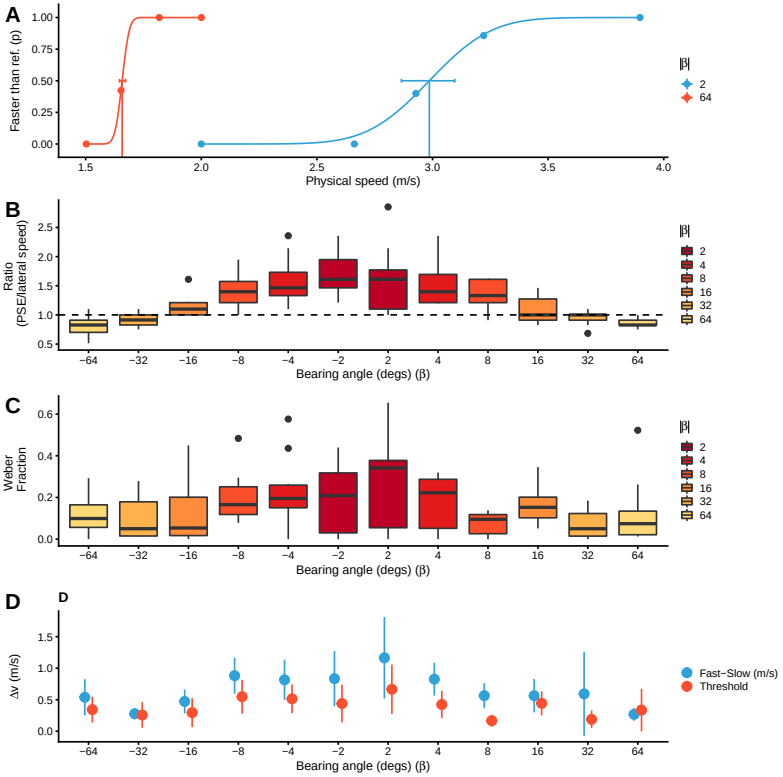

Figure 3.3. A) Psychometric functions for two different bearing angles for one participant (reference speed: 2 m/s). The y axis indicates the probability of judging the target's speed as faster than the reference speed. The horizontal error bar indicates the discrimination threshold. Speed in depth is underestimated with respect to the reference movement, as shown by the psychometric curve for β=2. Discrimination thresholds are higher for motion in depth compared to lateral motion. B) Average relative speed (PSE/Standard lateral). Values above the dashed line (Ratio > 1) denote underestimation of depth vs lateral speed. C) Weber Fraction as a function of β. Weber fractions are higher for motion in depth, indicating that observers are less precise when judging differences for MID compared to lateral movement. D) Representation of the differences between fast and slow trials (blue) and differential threshold (red) across β. Error bars indicate the 95% confidence interval. Mean differences for slow-fast trials are consistently higher than the discrimination threshold, therefore we assume speed was perceived as different.

suggesting that MID estimates follow Weber's law.

Finally, Figure 3.3D illustrates, for the sake of comparison, the speed differences between slow and fast trials and the measured differential thresholds. The speed difference (fast-slow) was significantly larger than the differential threshold $F(1, 198) = 24.63$, $p < .001$) and did not change significantly with β (i.e. interaction with β failed to reach significance, $F(11, 198) = 0.82$, $p = .620$). This provides sufficient grounds to assume that fast and slow speeds were perceived differently for further trajectory analysis.

3.5.2 Trajectory estimation

Figure 3.4 displays the average perceived trajectory across subjects and β split by slow/fast trials in a polar representation. Figure 3.5A shows the adjusted trajectory ($\hat{\beta}$) as a function of the physical trajectory of the stimulus (β).

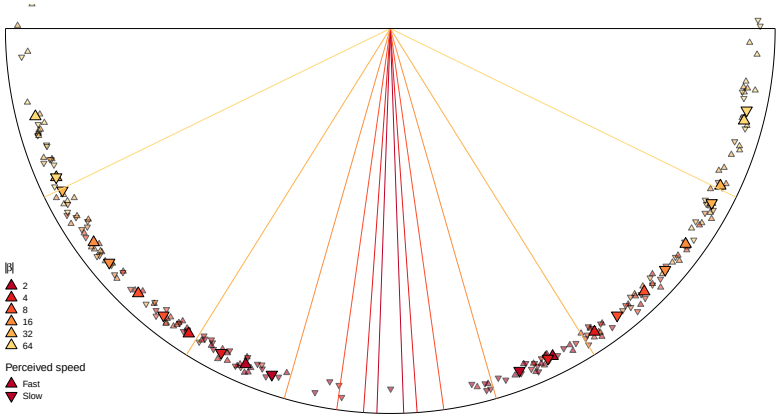

Figure 3.4. Polar representation of the average perceived trajectory pooled across subjects for each β and slow/fast perceived speed (upwards/downwards triangles indicate fast or slow group of trials respectively). Smaller triangles indicate observer's mean perceived trajectory split by β, perceived speed and participant. Jittering was added to the smaller triangles in order to ease interpretation.

An ANOVA on $\hat{\beta}$ yielded a significant (trivial though) effect of β ($F(1,$

3560) = 7495.39, $p < .001$) and, more importantly, speed group ($F(1, 3560) = 100.77$, $p < .001$) indicating that the perceived speed of the physical movement had an effect on the perceived trajectory and thus confirming our hypothesis showing that faster perceived speeds lead to more biased trajectory estimates. The interaction failed to reach significance ($F(1, 3560) = 2.44$, $p = .118$), indicating that the effect of the physical speed is independent of the presented trajectories in our study.

3.5.3 Bayesian model

As obtained by the optimization routine, the standard deviation of the slow prior is $V_\sigma = 0.329$. Figure 3.5B shows the performance of the Bayesian model split by reference speed and mean participant reports of the participants.

The predictions of the model strongly correlate with the reported estimates $\hat{\beta}$ ($r(3481) = 0.91$, $p < .001$). To check the suitability of our model predicting $\hat{\beta}$, we computed and compared the log likelihood of a model with real β and reference speed as a dependent variables against a model including only the predicted trajectories in the Bayesian model $\hat{\beta}_{\text{Pred}}$. The results show that the Bayesian model has a larger likelihood ($LogLik\ \hat{\beta}_{\text{Pred}}$: -15625.86) than the model based on β and reference speed ($LogLik\ \beta \times V_{Ref}$: -16414.7). This shows that our model predicts better $\hat{\beta}$ than a model relying on the physical trajectory (β) and reference speed.

3.5.4 Initial distance under-estimation

A Bayesian model can successfully explain several characteristics of this bias in the perceived trajectory (i.e. superior bias close to $\beta = 0$), participants variability (Welchman et al., 2008) and dependence with speed). However, a more parsimonious explanation for these results could be a simple misestimation of the initial distance on the target. Since viewing distance and target size are important variables to extract the rate of change in depth from the optic array, a constant underestimation of the initial distance would result in an overestimation of the trajectory of the target. We, therefore, checked if an underestimation of the initial distance could explain our results.

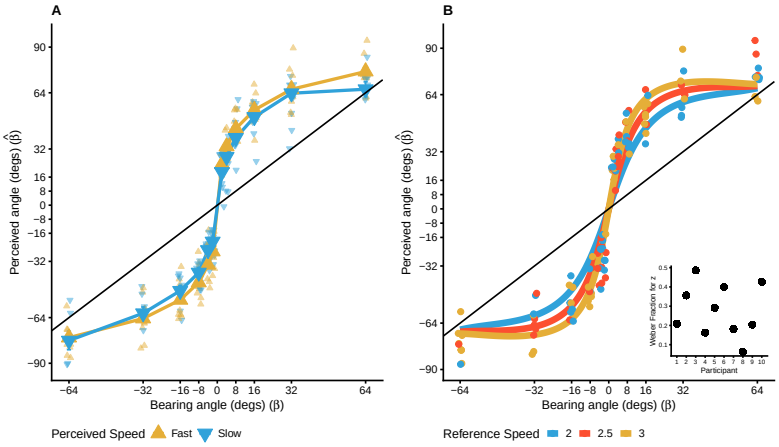

Figure 3.5. A) Average perceived trajectory between subjects as a function of the physical trajectory of the stimulus (β) split by fast/slow perceived speed (upwards/downwards triangles indicate fast or slow group of trials respectively). Error bar indicates the 95% confidence interval. Smaller triangles indicate observer's mean perceived trajectory split by β, perceived speed and participant. B) The curves represent the fit of the model across β split by reference speed. The points indicate the individual mean reported trajectories for each participant split by reference speed and β ($WF_x = 0.1$; $WF_z = 0.28$; $\sigma_v = 0.33$). Horizontal jittering was added to the individual points in both figures to ease interpretation. Inset indicates the Weber fractions used to estimate the perceived trajectories in the model for each participant.

In order to do so, we specified a linear model of the perceived trajectory ($\hat{\beta}$) as a function of the initial distance (D_0), and derived the expected value of the parameter that multiplies D_0 assuming there is no underestimation.

Given the following equivalence:

$$tan(\beta) = \frac{V_x}{V_z} \tag{3.22}$$

Multiplying V_x and V_z by the total time of presentation t_0 we obtain:

$$tan(\beta) = \frac{V_x \cdot t_0}{D_0} \tag{3.23}$$

Then, we can compute the value of ($V_x \ t_0$)) in our experiment and use it as a constant (k) in the linear model:

$$tan(\beta) = \frac{k}{D_0} \tag{3.24}$$

Therefore, if D_0 is underestimated we expect a slope larger than 1 dependent on the term $\frac{k}{D_0}$ since D_0 is in the denominator. Based on Eq. Equation 3.24, we applied a Linear Mixed Model to estimate this slope and check if it is larger than 1. Slopes of the ratio ($\frac{k}{D_0}$) were introduced as random effects, and the intercept was dropped since it is not present in Eq. Equation 3.24. The results for the fixed effects showed that the slope do not differ from 1 (Estimate = -6.479, $t(9.267)$ = -0.389, p = 0.706). Therefore, we did not find enough evidence to support the hypothesis that the initial viewing distance may have been underestimated.

3.6 Discussion

In this study we found a dependence between the speed and the perceived direction of an object, congruent with a Bayesian model of depth perception based on a slow prior as proposed by Welchman et al. (2008) and Stocker and Simoncelli (2006).

Our data show that the movement in depth is strongly underestimated with respect to lateral movement. A comparison between the Weber fractions for our trajectories (Figure 3.3C) shows that reliability depends on the z component, that, in turn, translates into the bearing angle (β). Consistent

with this, we have found that the variability in the estimation of the movement in depth is approximately two times greater. This is important, since, within a Bayesian framework depth estimates would be more affected than lateral estimates, presumably, causing the directional bias found in this and other studies.

In this study we were mainly interested in obtaining the PSE coupled with the estimated trajectory rather than obtaining a precise estimate of the discrimination threshold (ΔV). For this reason we chose the QUEST method. However, as Figure 3.3A ($\beta = 64$) shows, for a few conditions, the amount of speeds sampled by this procedure might affect these estimates. Because of this reason we had to formulate our Bayesian model assuming a Weber Fraction based on previous literature for the lateral movement. In addition, our results show that the PSE for $\beta = \pm 64$ is significantly lower than the reference speed ($t(18.0) = -5.16$, $p < .001$). This is against our first hypothesis, given that the rest of the angles show an underestimation of speed with respect to the reference. We hypothesized that, in this condition, for high speeds the target would have to move out of the projection frustrum before 1 s. (presentation time) possibly causing an effect on the perceived speed (Anstis & Kim, 2018). Consequently, the participants could have judged the test target as faster than the reference.

Our results are in line with previous literature in which the speed in depth was strongly underestimated with respect to lateral movements (Brenner et al., 1996; Brooks & Stone, 2006; Rushton & Duke, 2009; Welchman et al., 2008). For example, in Lages (2006), the participants were prompted to indicate both, the perceived trajectory angle and radial distance for stimuli moving in the $x - z$ space. Their results show a visual space describing an ellipsoidal shape as a result of V_z being underestimated with respect to V_x for the same physical speeds. Rokers et al. (2018) showed that a Bayesian model based on the slow motion prior can account for errors when judging the direction of an object. Their model is capable of predicting perceptual errors under different levels of contrast, eccentricity and distance to the stimulus. According to their results, the greater the distance to the stimulus, the greater the perceived lateral bias. The extent of trajectory bias in our study is higher than shown in the previous literature (J. Harris & Dean, 2003; Welchman et al., 2008). However, just as the

Rokers et al. (2018) model would predict, given that our initial distance is the highest (5, 1, and 0.5 m. respectively), the trajectory bias we would found would be the strongest.

Recently, Wei and Stocker (2017) have proposed a simple mathematical relationship that could be applied to this study. They hypothesize that perceptual bias is proportional to the rate of change of the discrimination threshold along the stimulus space, a notion that is supported by our results. At high angular speeds, lateral motion perception seems to follow Weber's law (Stocker & Simoncelli, 2006). Also, as we found in this study, Weber's law seems to apply to motion in depth. Therefore, the attraction towards the slow prior would be stronger for higher angular speeds (see Figure 3.1). This would account for an increased directional bias for higher speeds as shown by our results.

Even though our results seem clear, simulations of 3D movement in stereoscopic setups are known to introduce problems such as the conflict between vergence and accommodation. Conflict between these two cues could be responsible for consistent depth underestimation shifting perceived position closer to the screen due to accommodation for simulated depth movements (Regan et al., 1986), that is, increasing the bias in the perceived bearing angle. However, this conflict has been studied with contradictory outcomes for perceived distance, which according to our results would be a by-product of speed estimation: Watt et al. (2005) found that when accommodation was changed by manipulating the distance between display and observer, disparity scaling was corrected. Willemsen et al. (2008), however, found that vergence-accommodation conflict does not affect the perceived distance.

Future studies may explore the effect of a target moving with a non-null vertical component in order to constrain a full 3D motion model. It is known that sensitivities for stimulus oriented vertically or horizontally are nearly equal (Manning et al., 2018a; Portfors-Yeomans & Regan, 1996) and reports in the $x - y$ plane show little directional bias (Welchman et al., 2008). We thus expect our results to generalize to the $y - z$ plane. On the other hand, Poljac, Neggers, and Van Den Berg (2006) found that bias in the plane $y - z$ is lower than $x - z$, suggesting that vertical estimation may be less reliable than lateral estimation. Interestingly, Poljac et al.

(2006) found that direction was estimated more precisely when the target crossed the eye height. This suggests that additional visual cues may be used to estimate the direction of moving targets in this case. Furthermore, the participants performed the experiment in a free gaze situation. Since pursuit is known to influence the accuracy of speed estimates (Schütz et al., 2008), an interesting future direction would be to investigate the effect of smooth pursuit or fixation on the speed precision estimation for motion in depth and the resulting lateral bias.

Given that we have found a dependence between the estimation of speed and direction, there might be some common neural processes integrating both features. Lages and Heron (2010) previously proposed a parallel processing of 2D velocity estimates and disparity to extract estimates of 3D motion. 2D motion information would be encoded in V1 under a preference for slow speeds and selective directions (Perrone, 2006; Series et al., 2002). This could be interpreted as a prior for slow speeds (Vintch & Gardner, 2014). Then, this information is further processed by MT (Braddick et al., 2001; Burge & Geisler, 2015; Sanada & DeAngelis, 2014). Concerning disparity, an early computation of this signal is carried out by V1 (Nienborg et al., 2005). Further indirect projections to MT relay on V2 and V3 (Ponce et al., 2008) as an intermediate processing of disparity (Thomas et al., 2002). As a result, speed, direction and disparity processing are performed by MT. Therefore, it is the most probable candidate for the integration of these signals in order to obtain the structure of 3D motion perception (Rokers et al., 2009).

3.7 Conclusions

We show that the direction of an object on a near-collision course with the observer is overestimated as a function of the perceived speed. Objects are consistently judged as passing the observer further away than they actually do. Our methodology allowed us to couple the perceived speed for different bearing angles (β) with directional biases in depth perception. Our results indicate that a Bayesian model of speed discrimination in depth following Weber's law can successfully simulate both types of perceptual biases denoting a coherence between speed and direction estimation. Thus,

future research could manipulate the reliability of motion signals to further investigate the relation between motion and direction estimates.

4 Study 2: Gravity and known size calibrate visual information to time parabolic trajectories

Catching a ball in a parabolic flight is a complex task in which the time and area of interception are strongly coupled, making interception possible for a short period. Although this makes the estimation of time-to-contact (TTC) from visual information in parabolic trajectories very useful, previous attempts to explain our precision in interceptive tasks circumvent the need to estimate TTC to guide our action. Obtaining TTC from optical variables alone in parabolic trajectories would imply very complex transformations from 2D retinal images to a 3D layout. We propose based on previous work and show by using simulations that exploiting prior distributions of gravity and known physical size makes these transformations much simpler, enabling predictive capacities from minimal early visual information. Optical information is inherently ambiguous, and therefore, it is necessary to explain how these prior distributions generate predictions. Here is where the role of prior information comes into play: it could help to interpret and calibrate visual information to yield meaningful predictions of the remaining TTC. The objective of this work is: (1) to describe the primary sources of information available to the observer in parabolic trajectories; (2) unveil how prior information can be used to disambiguate the sources of visual information within a Bayesian encoding-decoding framework; (3) show that such predictions might be robust against complex dynamic environments and (4) indicate future lines of research to scrutinize the role of prior knowledge calibrating visual information and prediction for action control.

This study has been published as: Aguado, B., & López-Moliner, J. (2021a). Gravity and known size calibrate visual information to time parabolic trajectories. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.642025

4.1 Introduction

Intercepting a ball in a parabolic trajectory before reaching ground level is a fundamental task in different sports: batting a baseball, hitting a high lob in tennis, or heading a football. In those situations, the time at which the interception is possible is very tight, yet our performance is astonishing. Time-to-contact (from now on TTC), that is, the time until an object reaches a location of interest, can provide very useful information that would help anticipate motor programs to solve those tasks.

In principle, to intercept a target, it would be enough to estimate its position and predict its future position based on speed estimates. Solutions based on this idea have been put forward for 2D motion (Aguilar-Lleyda et al., 2018; Kwon et al., 2015) but the generalization to 3D parabolic trajectories faces complex problems deeply rooted in the inverse-projection problem of Perception. The inverse problem of Perception refers to the ambiguous mapping between a distal stimuli and final percept (Kersten et al., 2004; Pizlo, 2001). Unlike previous attempts where TTC is obtained from optical variables, in this paper, we propose that some constants in the environment like gravity and size are considered and ease the otherwise complex transformation of optical variables to a 3D world to obtain relevant variables like TTC. The stance taken in this work will assume that we make implicit inferences (Helmholtz, 1867) about the present and future states of the world to act. However, the nature of the information guiding the control of action is an ongoing source of debate within the study of Perception.

4.1.1 Two theories for interceptive control

Information-based control

The information-based control perspective, rooted in the Ecological or Gibsonian framework of Psychology (Gibson, 1966, 1979) (Figure 4.1A), assumes that perceptual information is governed by certain physical regularities (Turvey et al., 1981) that can be captured and exploited to control our action. Under ecologically valid conditions (i.e. full-cue conditions), our perceptual system would be attuned to perceptual invariants directly specifying the characteristics of an event without the need to perform in-

ternal computations according to which humans act (Gibson, 1979). Thus, the information to solve a given task is directly specified in the optic flow (direct perception) explaining why only identifying task-relevant visual variables will determine an actor's successful action from a perceptual perspective.

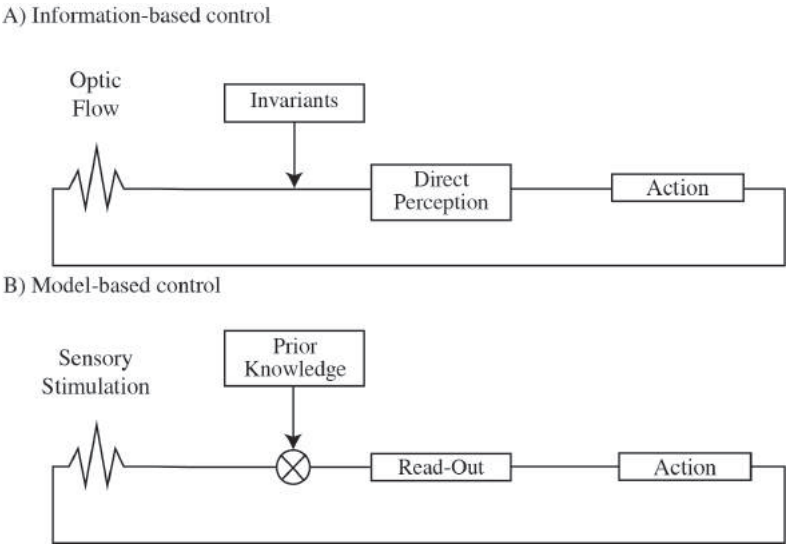

Figure 4.1. A) Optic flow conforms to invariants that specify properties of the environment (direct perception), indicating the adequacy and availability of action within the task. B) Sensory stimulation is combined with prior information to infer current or future states in the environment (read-out), providing the grounds to plan and adapt action.

Under this framework, mainstream interpretations argue that the role of the observer is to actively seek out invariants within certain task-relevant pieces of optic information and unfold a coupled action based on instantaneous information. Following this line, detecting and maintaining invariant stimulation requires reducing the difference or the error with an "ideal value". Because of that, these strategies are also called error-nulling strategies (Fajen, 2005b). Based on this idea, different control laws have been proposed for visually guided actions such as intercepting a moving object (Bruggeman et al., 2007; Warren et al., 2001; Wilkie & Wann, 2003; Zhao et al., 2019), braking (Lee, 1976) or catching a ball on the fly (Chap-

man, 1968; McBeath et al., 1995; McLeod & Dienes, 1993; Michaels & Oudejans, 1992).

Among the previous examples, catching a ball in a parabolic flight is a paradigmatic interception case including locomotion and manual interceptive phases. The study of the underlying mechanisms regarding the locomotive phase is usually referred to as the outfielder problem (Todd, 1981). The outfielder problem studies a case of interception in which baseball players known as outfielders must move to catch a high-flying ball in a parabolic trajectory before it hits the ground (Chapman, 1968; Michaels & Oudejans, 1992; Todd, 1981).

The first catching error-nulling strategy put forward to explain action control within the outfielder problem was Chapman's strategy (Chapman, 1968). S. Chapman noticed that when a ball follows a parabolic trajectory on a collision course with the observer, the elevation angle (γ) (see Figure 4.2), that is, the vertical angle between the ball's position and an observer's eye level, increases during the whole trajectory. Therefore, all an observer would need to do to navigate towards the interception location is keeping the elevation angle increasing during the whole trajectory at a constant rate. However, others suggested it should increase at a decreasing rate (McLeod & Dienes, 1993, 1996).

Later, Michaels and Oudejans (1992) described Chapman's strategy in terms of the projected image in the vertical plane at launch distance. If the ball is in a collision course with the viewer, the vertical projection of the ball increases linearly through the trajectory. In any other case, the image of the ball would displace non-linearly, that is, accelerated. Therefore, to catch a ball in a parabolic trajectory, one needs to actively maintain the acceleration of the projected vertical position of the ball at zero. Motivated by that, this strategy was named Optic Acceleration Cancellation (from now on, OAC) in McBeath et al., 1995.

Although the error nulling strategies emerged within the ecological framework, they conflict with a key concept at the core of the Ecological theory, the theory of affordances (Gibson, 1979). The affordance-based theory emphasizes the idea that observers are tuned to the availability of an action given a sensory array. This tuning would be a by-product of a gauging process that maps optic into movement information and even into

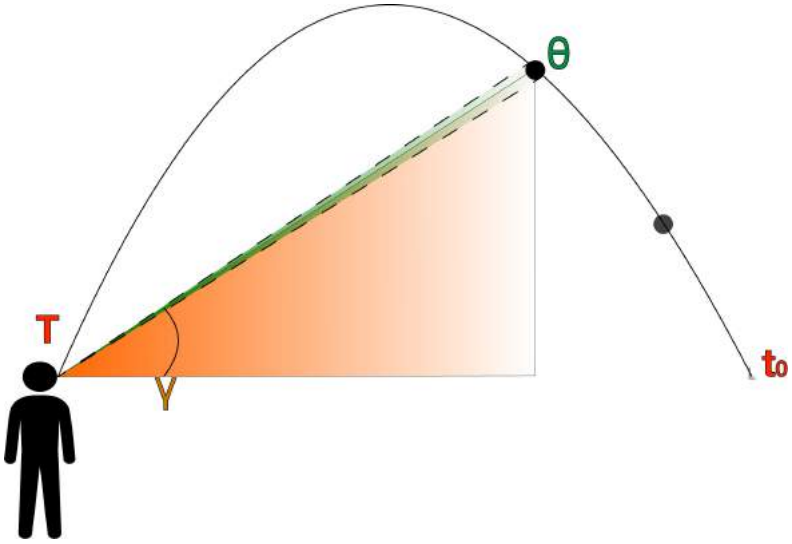

Figure 4.2. Lateral view of a parabolic trajectory depicting the primary primitive monocular cues, that is, retinal size (green projection; θ) and elevation angle (orange projection; γ).

optic correlates in object size units (Jacobs & Michaels, 2006; Peper et al., 1994). Hence, if a fielder is correctly calibrated, acting to keep certain variables of interest into a "safe region" would ensure interception. This notion provides the grounds for an affordance-based control strategy theory (Fajen, 2007).

Reformulating the OAC under the scope of the affordance-based control, either cancelling out vertical acceleration or running at maximum speed without being able to cancel out vertical acceleration, would be required to perceive catchability. However, a series of studies (Postma et al., 2017; Postma et al., 2018) found that actual catchers did not need to cancel out acceleration nor run at their maximum speed to judge catchability. These results cast doubts on the informational nature of a catchability affordance.

Despite a lesser dependence on immediate visual information, an affordance-based control strategy is still dependent on instantaneous visual information. In this respect, simulations of the locomotion based on error nulling strategies showed irreconcilable with actual catches when

accounting for human neuromotor acceleration and sensorimotor integration delays (Kistemaker et al., 2009; Kistemaker et al., 2006; McLeod & Dienes, 1996). Consequently, a minimal prediction seems necessary to account for sensorimotor delays in the central nervous system (Nijhawan, 1994). Furthermore, the occlusion of visual information would result in a considerable impairment of the action. A possible solution to provide adequate responses would be to continue to do what has been done so far (Bootsma, 2009). Nevertheless, in temporally constrained tasks, another question arises: for how long? In the outfielder problem, catching the ball in-flight imposes tight temporal restrictions constraining the interception area. Hence, predicting parameters such as TTC or the interception area in dynamic contexts is key to planning our actions while looking around for a teammate, finding a safe path towards the goal, or modulating speed to reach the interception location in time. Following this reasoning, some alternatives have been proposed in the literature as solutions based on predictions about future states of the world.

Model-based control

The model-based control is framed within the constructivist framework (Helmholtz, 1867). It assumes that the information picked up by our senses is inherently ambiguous, to some degree corrupted by noise in the neural system and delayed at higher-order brain areas. (Craik, 1967) proposed that the brain tries to infer and replicate an external world model given the available sensory information. This replica results in an internal model of the environment, including an agent's state that allows one to predict future states of the world and act accordingly avoiding sensorimotor delays.

Relying on predictions would allow us to divert our gaze from the immediate region of interest and consequently interrupt the sensory flow. For example, Hayhoe et al., 2005 and Diaz et al., 2013 showed anticipatory saccades towards future interest points to plan future goal states. The same applies to catching a ball in a parabolic trajectory for manual interception tasks. Despite distractors, parallel tasks, occlusions or head-turns that might divert our attention, we still manage to intercept a ball in flight (Binaee & Diaz, 2019; Dessing et al., 2009; López-Moliner & Brenner, 2016; López-Moliner et al., 2010). In fact, we can hit it even when the ball

was visible for just a short time (Amazeen et al., 1999; Sharp & Whiting, 1974; Whiting & Sharp, 1974), revealing that actually, the major constraint for the use of a predictive strategy would be to obtain predictions early enough to overcome sensory-motor delays.

The trajectory prediction strategy (Saxberg, 1987a, 1987b), framed within the model-based theory, assumes that an observer predicts where and when the ball will be within reach in Cartesian units, allowing to pre-program a minimal action plan since motion onset. However, the available optic information is egocentric and therefore ambiguous with respect to its source. Therefore, an observer must perform an inferential process to interpret optic information accurately.

In this line, Perception has been proposed as a Bayesian inferential process in which visual information is interpreted as a function of the most probable state of the world given prior knowledge. This inferential process has been formulated in terms of "encoding" and "decoding" (Friston, 2010; Knill & Pouget, 2004; Wei & Stocker, 2015). Encoding corresponds to the activity resulting from the transduction of external energy onto the sensory receptors. The encoded sensory information is then combined with prior knowledge through an inferential process called decoding (Figure 4.1B). The product of the decoding is an interpretation (read-out) of the currently available data resulting in a belief of the state of the world that provides the grounds to draw predictions.

In real life, we generally do not judge the parameters of a task for ambiguous targets in the environment. Instead, we have some prior knowledge of the elements to be judged that are stable and might help disambiguate optic information providing the grounds to extract valuable information for the task. In this line, a significant number of works highlighted the role of prior knowledge about contextual variables such as gravitational acceleration (Jörges & López-Moliner, 2017; McIntyre et al., 2001) or known/familiar size (Hosking & Crassini, 2010; López-Moliner et al., 2007), framing the interpretation of visual information for the control of timed actions.

Note that using a priori knowledge does not imply the availability of accurate Cartesian metrics or Newtonian laws within an internal model. A fully-featured 3D internal model replicating the external world has been

repeatedly dismissed. For example, Shaffer and McBeath, 2005 showed that even expert baseball players could not judge the apex of a ballistic trajectory on a collision course with the observer. In this situation, the apex was estimated to be 0.33 seconds before collision for flight durations of 4 seconds, that is, 1.66 seconds after the actual apex. These results indicate a tendency to judge the apex of the elevation angle as the physical apex of the trajectory. Also, Reed et al., 2010 showed that neither expert nor novel baseball players could reconstruct the visual trajectory of a parabolic trajectory in a head-on approach mixing up the ball's movement in space with the visual trajectory it follows.

Unlike Craik, 1967, we propose that the prior knowledge can be kept to a minimal number of components that help exploit the optic flow's complexity. Thus, in line with a Bayesian framework, under our view, the use of prior knowledge just suggests the existence of a probabilistic and implicit knowledge acquired by repeated experience that helps infer the most probable sources of visual information in the external world (Gómez & López-Moliner, 2013; Zago et al., 2009). In this sense, an accurate representation of a priori parameters would suffice to obtain reliable estimates of the task's parameters.

Here we propose using priors as internalized knowledge to translate optic variables into temporal estimates in a process we name calibration. Calibration would be the process by which optical or angular information is mapped into Cartesian ones with the assistance of different pieces of prior knowledge providing actionable predictions (López-Moliner et al., 2007; López-Moliner et al., 2013). Calibrating optic cues into Cartesian allows us to test the correspondence between the prediction of a model and our actions in terms of accuracy. Furthermore, it may allow us to formulate hypothesis based on known psychophysical precision levels for the different pieces of information in the sensory array and even check if integration rules apply (de la Malla & López-Moliner, 2015; Wolpert et al., 1995).

As an example, the GS model (Gómez & López-Moliner, 2013) is an algorithm that predicts the TTC for parabolic trajectories based on a combination of optic variables and prior knowledge information. Its predictions have been partially validated based on predictions about the accuracy and the precision of temporal estimations (Aguado & López-

Moliner, 2021b; de la Malla & López-Moliner, 2015). However, there is a lack of mathematical formulations to predict the interception location. This makes it harder to test experimentally predictive control strategies within the outfielder problem. Hence, in this work, we will limit ourselves to indicate the role of TTC estimates guiding the interceptive action.

It is essential to mention here that our definition of calibration is different from the definition of calibration made within the Ecologic framework (Fajen, 2005a; Jacobs & Michaels, 2007). Under our perspective, calibration is a process by which otherwise ambiguous optical information is directly mapped into kinematic and temporal estimates such as motion vectors or TTC highlighting the relevance of prior knowledge to provide predictions that may assist visually guided actions. In contrast, within the Ecological framework, calibration would be a by-product of a gauging process that maps visual information into movement information and even into optic correlates in object size units (Jacobs & Michaels, 2006; Peper et al., 1994).

Nevertheless, producing predictions does not necessarily mean that those predictions will be accurate or that the new visual information would be disregarded. Take the case of Fink et al., 2009 study. Participants had to catch a ball in a parabolic trajectory that suddenly would alter its motion towards the ground. As a reaction, the catchers changed their trajectory towards the ball as well which was taken as support for the information-based control perspective. Under Fink et al., 2009 rationale, a model-based control strategy would result in a consistent path towards the interception point despite mid-flight disturbances. This rationale assumes that new information would be dismissed or might be irrelevant because the prediction would remain invariant. However, predictions would also be subject to continuous evaluations to avoid errors or perceptually driven biases. In a similar line, Postma et al., 2014 reasoned that continuously gazing the ball through the trajectory would support the information-based control perspective. However, following the ball with our gaze does not necessarily imply that action guidance must be driven by instantaneous optic information. Periodically sampling visual input to correct the prediction made would be an alternative strategy to guide action with reasonable levels of accuracy in a more general framework (Brenner & Smeets, 2018). In fact, a simulation

study conducted by Belousov et al., 2016 showed that predictive behaviour would be indistinguishable from using error-nulling strategies if the ball is continuously monitored.

In the following sections, we will unveil how the interaction between perceptual information and prior knowledge contributes to interpreting and reliably predict TTC for gravitationally accelerated objects under parabolic trajectories. To do so, we will first analyse the available sources of visual information within the optic flow to judge TTC or time an interceptive action. Then, we will stress the role of prior information in calibrating visual information. After that, we will show the accuracy and reliability of the GS model which includes known gravity and size in complex environments. Finally, we will indicate future lines of research to address the role of predictions of TTC guiding interceptive behaviour.

4.2 Available visual information

When an observer faces a ball in a parabolic trajectory, the projectile describes the following sequence of events (see Figure 4.2). The ball initially goes up at a decreasing speed until it reaches the peak of its trajectory. Then, it accelerates during the descent towards the ground. However, an observer cannot access the underlying dynamics of projectile motion using Cartesian metrics. Instead, they only have access to information based on egocentric angular variables that depend on their position and the kinematics of the ball (McBeath et al., 2018; Shaffer & McBeath, 2005). Hence, the ball's kinematics and an observer's movement influence the visual information being exposed throughout the trajectory. Figure 4.2 depicts the main primitive optic variables that we will consider and how both unfold over time (t) depending on the observer's position and total flight time (T), which is unknown to the observer.

While interpreting sensory information is a central part of this work, we also need to consider the limitations of our sensory system to gather that visual information. Because of that, in the following subsections, we will elaborate on the conditions that render different visual cues helpful regarding detectability or discriminability, describing their precision as Weber fractions. To do so, we will assume that the observers keep their

gaze on the projectile, which is usually the case in free viewing situations (Oudejans et al., 1999; Postma et al., 2014) and laboratory-controlled tasks when the position to hit a target is not pre-specified (Brenner & Smeets, 2007; Cámara et al., 2018; Soechting et al., 2009).

4.2.1 Retinal size (θ)

The visual angle or retinal size is the angular size projected by an object on the retina (θ; see green projections in Figure 4.2). Retinal size (θ) is proportional to both object size and distance, being the prototypical example of an ambiguous optic variable. Previous studies have shown that human performance in discrimination tasks for angular size judgments is about a 3 - 6% Weber fraction (WF) (k_θ) for objects yielding > 0.0009 radians, increasing steeply up to a 20 - 30% (WF) for smaller objects (Klein & Levi, 1987; McKee & Welch, 1992; Westheimer & McKee, 1977).

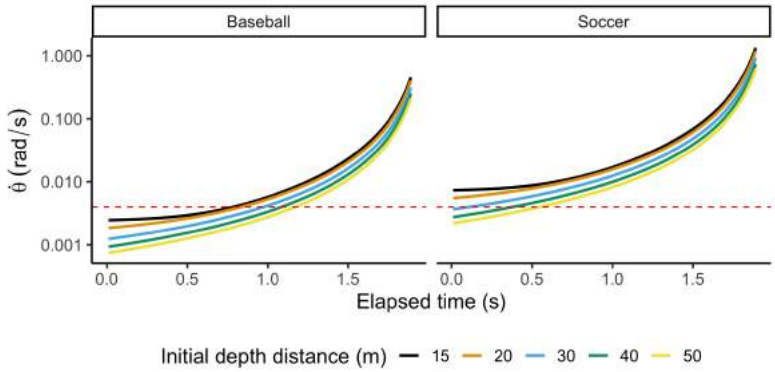

Figure 4.3. Rate of expansion ($\dot{\theta}$) for two different ball sizes at five different initial distances. The total flight time is two seconds. Values under the red dashed line (0.004 rad/s) indicate that an observer cannot discriminate differences.

Retinal size is a zero-order variable; that is, it does not carry temporal information and thus, cannot be employed alone to estimate motion components or TTC. To do so, one needs to have access to the rate of expansion ($\dot{\theta}$), which is the speed at which the retinal image changes. The absolute detection threshold for $\dot{\theta}$ has been reported to be about 0.0003 rad/s (McKee,

1981), while the discrimination threshold associated with this parameter is about an 8.5 - 14% of change (WF) ($k_{\dot{\theta}}$) (Regan & Hamstra, 1993).

In some cases, such as baseball games, the players meet scenarios where the ball is at a considerable distance. In those cases, although mediated by physical size and ball's horizontal speed, discriminability of the rate of expansion ($\dot{\theta}$) is generally poor (20% of change (WF) for objects expanding at a rate of less than 0.004 rad/s (J. Harris & Watamaniuk, 1995; Regan & Beverley, 1978).

To show the effect of ball size on an observer's ability to discriminate differences in the rate of expansion ($\dot{\theta}$), we computed the rate of expansion for two different balls (baseball and soccer balls) moving in a parabolic trajectory towards an observer from different initial distances. Figure 4.3B shows how retinal expansion unfolds as a function of time. Values below 0.004 rad/s (red dashed line) would fall below the optimal discriminability range, indicating that the observer's ability to discriminate differences is inferior. As depicted in Figure 4.3B, a player facing a baseball will not discriminate retinal expansion during more than half of the trajectory. Instead, differences in retinal expansion can be discriminated during most of the trajectory of the Soccer ball. This example aims to point out that even when visual cues are present in the optic flow, the resolution of our visual system might not allow us to use them to guide our actions. Therefore, initial estimates of TTC might be computed using alternative routes.

4.2.2 Tau (τ)

Lee's seminal work (Lee, 1976) described Tau (from now on, τ) as the ratio between the visual angle (θ) and its rate of expansion ($\dot{\theta}$) (see Hecht and Savelsbergh, 2004 for a review). Tau signals TTC and is directly accessible within the optic flow without prior knowledge or previous estimates of distance, size, or approaching speed.

Although τ can not be conceptualized as a primitive variable in the study of visual cues, it has shown different features that could allow us to consider it as such. Regan and Hamstra, 1993 found that differences in τ could be distinguished independently of differences in θ or $\dot{\theta}$ ($k_{\dot{\tau}} \approx$ 0.07:0.13; WF). Because of this, Regan and Hamstra, 1993 concluded that there might exist a mechanism sensitive to τ independently of θ or $\dot{\theta}$.

Indeed, other studies have shown the existence of a neural mechanism tuned to τ (independently of retinal size and rate of expansion) or some of its modifications (such as the η-function Judge and Rind (1997) or τ_m-function Keil and López-Moliner (2012)) in various species such as pigeons and humans (Rind & Simmons, 1999; Sun & Frost, 1998; Yonas et al., 1977).

Tau (τ) has been indicated as a source of prospective information that might be used as a threshold (as a criterion) to perform different tasks such as hitting (Bootsma & van Wieringen, 1990; Lee et al., 1983) or catching (Savelsbergh et al., 1991). This threshold is usually referred to as Tau-margin (Wann, 1996). However, its applicability to time parabolic trajectories might be compromised due to several limitations.

First, τ would only generate accurate TTC predictions at launch when the ball travels in a collision course with the observer.

Second, the object should approach the observer at a constant speed. In a parabolic trajectory, from an allocentric viewpoint, V_x and V_z are constant (assuming no air resistance). Nevertheless, the approaching speed for the observer corresponds to radial velocity (V_r), which would carry the isotropic expansion of the retinal expansion. However, V_r is not constant through the trajectory and cannot be directly estimated from the optic flow (Gómez & López-Moliner, 2013). For instance, consider a trajectory launched 2 meters far from an observer in a trajectory 10 meters of height. In this situation, during the first half of the trajectory,V_r and retinal expansion $\dot{\theta}$ would be negative; that is, the object moves further away from the observer, rendering meaningless estimates of TTC using Tau.

Third, even though Tau could be discriminated independently of the rate of expansion ($\dot{\theta}$), it is likely constrained to the same detection thresholds (Keil & López-Moliner, 2012). Therefore, rates of expansion ($\dot{\theta}$) lower than 0.004 rad/s could result in a non-informative source to guide temporal estimations during an important section of parabolic trajectories. As reported above, the Weber Fraction for Tau ranges between 7% to 13%. Therefore, we will be using a mean Weber fraction of 10% referring to Tau in the following sections.

Finally, it might not be directly implemented as a general-purpose mechanism because the object must be spherical and rigid, which is not the case for interception in some sports such as Rugby or Frisbee.

4.2.3 Elevation angle (γ)

The elevation angle (γ) is the position of an object in the vertical meridian of the retina (see orange projection in Figure 4.2). We conceptualize the elevation angle (γ) using a spherical projection because angular variables are assumed to be directly accessible to an observer (McBeath et al., 2018).

Even though it is often referred to as the vertical position of an object in the retina, when a projectile is visible, we tend to perform continuous visual follow-ups foveating the object, which might prevent perceptual biases (de la Malla et al., 2017). In those cases, the retinal angle of the elevation angle (γ) would tend to zero. Because of that, it has been suggested that this visual angle can be estimated as a combination of the displacement of the environment in the retina (Brenner & Smeets, 2015; Oudejans et al., 1999), the movement of the eyes with respect to the observation axis (Crowell & Banks, 1996) and estimates of the heading angle generated at the otoliths of the vestibular system (Berthoz, 2000; Roy & Cullen, 2003) produced by the movement of the head and trunk (Crowell et al., 1998; Lewis et al., 1998). Previous literature has found that foveated objects require a difference of up to a 3 - 5% of change (WF) to be effectively discriminated (Crowell et al., 1998; Regan & Kaushal, 1994).

The first derivative with the time of the rate of change of the elevation angle ($\dot{\gamma}$) is the vertical rate of displacement of a target in the retina (see Figure 2). According to several studies (de Bruyn & Orban, 1988; Orban et al., 1984), the ability to judge differences in $\dot{\gamma}$ is about 5% (WF) for angular velocities between 0.03 rad/s (1.71 deg/s) and 1.2 rad/s (69 deg/s). Interestingly, Portfors-Yeomans and Regan, 1996 suggest channels that process position and cardinal motion independently, which indicates that the noise for both estimates is independent.

Given that parabolic trajectories move accelerated by terrestrial gravity, it is reasonable to consider humans' ability to detect acceleration. Calderone and Kaiser, 1989 proposed that acceleration in the visual system can be studied as the rate of change in speed divided by the average object speed in a two-stage process carried out in about 200 ms. (Werkhoven et al., 1992; Zaal & Bootsma, 2011). This delay would mean that the observer would not continuously monitor the adequacy of their actions. Furthermore, some studies found that it is necessary at least $\tilde{2}0\%$ of the change in speed to

detect acceleration (Babler & Dannemiller, 1993; Brouwer et al., 2006; Gottsdanker et al., 1961b; Werkhoven et al., 1992; Zaal & Bootsma, 2011) indicating that humans are quite insensitive to changes in speed.

4.2.4 Disparity (δ)

Disparity (δ) is the vertical/horizontal difference between the position of an object in the retinal image of both eyes. An algorithm like Tau (τ) for the estimation of TTC has been proposed as a combination of the knowledge of interocular distance (I), the distance with the ball (D) and the rate of change of horizontal disparity ($\dot{\gamma}$) (TTC $\approx \frac{I}{D*\dot{\gamma}}$). Furthermore, it can be used to estimate the lateral distance at which an object would pass an observer position.

A combination of Tau and the information contained in above expression would assist the estimation of TTC to achieve our exceptional temporal precision batting fastballs (Gray & Regan, 1998). This solution may account for systematic underestimations of TTC by weighting the visual cue that indicates a shorter TTC to guide the final interceptive phase (Gray & Regan, 1998; Rushton & Wann, 1999; Savelsbergh & Whiting, 1992). However, Brenner et al., 2014 found no evidence that hitting a free fall ball uses the rate of change in disparity ($\dot{\gamma}$) to estimate TTC. In this sense, Brenner and Smeets, 2018 argue that some studies that compare the performance between monocular and binocular conditions "ignore the benefit of having two estimates of the relevant monocular cues" instead of one.

4.3 Evidence of prior knowledge calibrating visual information

Humans quickly acquire knowledge about regularities in their interaction with the environment. Regularities such as the light coming from above (Adams et al., 2004), that bigger means heavier (Peters et al., 2016) or the fact that object size is generally constant (López-Moliner & Keil, 2012) enhance predictability and reduce uncertainty about future states of the world. Indeed, the assumption of a stable world is at the heart of essential findings in different areas of vision science, such as speed perception

(Stocker & Simoncelli, 2006), depth estimation (Glennerster et al., 2006) and manual interception (Brenner & Smeets, 2015).

Such known regularities may also be referred to here as contextual information stressing the role of acquired knowledge by repeated experience within a specific context. In this work, we will focus on two pieces of internalized knowledge that usually remain stable in our world and might frame the interpretation of visual information for parabolic trajectories: known size and gravity.

4.3.1 Size

The assumption of constant size is likely one of the most critical assumptions about action because, in general, the objects around us do not change size unexpectedly. Under this assumption, known size calibrates visual information into distance estimations with an object as the ratio between known size and the retinal size projected (Hecht et al., 1996; Ittelson, 1951; Sousa et al., 2011) even though irregular objects such as rugby balls and frisbees can be problematic.

$$d \approx s/\theta \tag{4.1}$$

The previous expression provides a mechanism to scale the optic space into ball size units (Gómez & López-Moliner, 2013; Peper et al., 1994) for a broad range of contexts. Calibrating the optic space with known size provides estimates of relative distances in paintings, pictures, video games, and environments with low or incongruent pictorial detail (Saxberg, 1987b; Todd, 1981), sometimes at the cost of leading to systematic misperceptions (Battaglia et al., 2011; Battaglia et al., 2005; Tcheang et al., 2005). In line with the use of known size calibrating visual information, López-Moliner et al., 2007 showed that when an object approaches at a constant speed and physical size is known; an observer can exploit the lawful relations between physical size and optic variables in the equation above to estimate approaching speed (V_z) as its first derivative:

$$V_z \approx \frac{s \cdot \dot{\theta}}{\theta^2} \tag{4.2}$$

In the above expression, known size (s) allows calibrating retinal size

(θ) and rate of expansion ($\dot{\theta}$) into an estimate of approaching speed (V_z) from otherwise spatio-temporally ambiguous optic variables. Then, an observer could use single optic variables (e.g. retinal size or expansion rate) to time the initiation of interceptive actions (López-Moliner et al., 2007; López-Moliner & Keil, 2012; Smith et al., 2001) depending on noise levels (Aguilar-Lleyda et al., 2018). However, what happens when it comes to obtaining parameters of more complex tasks such as parabolic trajectories?

In Todd, 1981, the participants had to judge if a ball on a parabolic trajectory would fall in front or behind an observer in different experimental conditions. Todd's work showed that, even though there might be enough information to estimate the final position qualitatively for parabolic trajectories only based on sensory information, prior knowledge of an object's size helped the participants to judge the final position in depth accurately. In each block, the absolute size was either fixed, selected at random, or fixed to a single dot during the whole trajectory. Accuracy was significantly better when the absolute size was fixed than the condition in which size was selected randomly. Those results indicate that prior knowledge of the ball's size aids the estimation of motion-in-depth.

Interestingly, the third condition yielded the worst performance of all three conditions, yet performance was slightly over chance level. In this condition, only the vertical movement was available to the observers to judge approaching speed. Participants judged landing position based on "the amount of vertical speed." In line with these results, Jörges and López-Moliner, 2017 showed that prior knowledge of gravity might be essential to calibrate estimates of the rate of change ($\dot{\gamma}$) into estimates of approaching speed (V_z) in parabolic motion.

4.3.2 Gravity

Since Lacquaniti and Maioli, 1989 work, showing an anticipatory activity for gravitationally accelerated objects, there is evidence that an internal representation of gravity may play a key role in controlling interceptive actions and judging TTC. For example, McIntyre et al., 2001 found that astronauts react to moving objects as if they were accelerated by Earth gravity under micro-gravity conditions. That study showed that although the astronauts were immersed in an environment where visual and bodily cues

indicated microgravity conditions, they could not adapt their interceptive actions completely. After 15 days, the astronauts were still anticipating their reaching, mimicking the conditions under terrestrial gravity. Subsequent work using virtual reality setups showed that participants could adapt to arbitrary gravities in a few trials. However, the performance is still lower than that under terrestrial gravity conditions (Zago et al., 2005; Zago & Lacquaniti, 2005).

Since then, an implicit representation of gravity has been found at a neurobiological level (Indovina, 2005; Miller et al., 2008) and in a broad range of tasks such as eye behaviour (Bosco et al., 2012; Diaz et al., 2013; Jörges & López-Moliner, 2019) or the estimation of the duration of events (Hosking & Crassini, 2010; Jörges et al., 2021; Moscatelli & Lacquaniti, 2011) despite our general insensibility to accelerations (Werkhoven et al., 1992).

However, the best example of a representation of gravity for sensorimotor control is that an observer does not need to see an ascending ball falling to intercept it (de la Malla & López-Moliner, 2015). In general, humans have an implicit expectation that upwards moving objects will eventually fall (López-Moliner et al., 2010; Reed et al., 2010). Nevertheless, this representation may not be available for every kind of tasks. For example, timing tasks for gravitationally accelerated objects in imagination show a bias towards the last visible motion speed (Bratzke & Ulrich, 2021; Gravano et al., 2017). In this line, some authors dismiss an internal model-based explanation favouring a prediction-free explanation (Baurès et al., 2007; Katsumata & Russell, 2012). However, the lack of adaptation under microgravity conditions and the need to account for sensorimotor delays pinpoint the relevance of a gravity prior guiding predictive control (Zago et al., 2008).

A recent study (Jörges & López-Moliner, 2020b) tried to derive the mean and standard deviation of the Gravity prior in a Bayesian framework. Their results found a prior with a standard deviation of 14% (WF). According to the authors, these results might correspond with an upper bound, as there seem to be theoretical reasons such as the lack of adaptation to arbitrary gravity values suggesting a relatively inflexible and robust gravity prior (Jörges & López-Moliner, 2017).

To test the use of different pieces of prior information for calibration, it is first necessary to put forward algorithms that require pieces of internalized knowledge. In the temporal domain, (Gómez & López-Moliner, 2013) showed that by using both prior knowledge of gravity and size, visual information could be calibrated, resulting in actionable estimates of TTC. The model was named GS model about the assumption of a priori known gravity and size.

4.4 Time-to-contact estimation

4.4.1 GS model

The GS model (Gómez & López-Moliner, 2013) is an algorithm that relies on calibrated optic information using prior knowledge to obtain estimates of TTC for parabolic trajectories. It relies on a combination of contextual variables such as known ball size (s) and gravitational acceleration (g) along with monocular cues such as retinal size (θ), elevation angle (γ) and its first derivative ($\dot{\gamma}$), providing accurate estimates of TTC.

$$TTC_{GS} = \frac{2}{g}\,\frac{s}{\theta}\,\frac{\dot{\gamma}}{cos(\gamma)} \tag{4.3}$$

Known ball size (s) and retinal size (θ) provide a mapping from retinal to Cartesian metrics. On its part, gravitational acceleration (g) calibrate and normalize the rate of change of the elevation angle ($\dot{\gamma}$) to be interpreted into meaningful predictions of TTC under arbitrary gravitational accelerations. In addition, cos(γ) would act as a non-declarative internalized parameter linked to action expecting that the elevation angle (γ) would increase over time (Reed et al., 2010; Shaffer et al., 2013). Removing the internalized variables from the GS model, one can still obtain a correlate of TTC based on retinal size (θ), the elevation angle (γ) and the rate of change of the elevation angle ($\dot{\gamma}$). However, its value is meaningless in signalling an actionable TTC and, therefore, not directly applicable. In the following section will illustrate that envisioning those pieces of prior knowledge as priors within an encoding-decoding framework could calibrate ambiguous visual cues into accurate estimates of TTC for parabolic trajectories.

Simulating the benefits of using gravity and size priors for the decoding

Under the constructivist framework, visual information is underspecified, and many trajectories can originate the same stimulation. This simulation shows how entering the correct gravity and size values increases the chances of inferring the actual trajectory exposed to the system from a subset of the possible ones.

In the case of this simulation, the possible inferred parabolas are a combination of the nine different TTC (ranging from 1.8 to 2.2 s in steps of 0.05 s), five different conditions of gravity (8.826, 9.316, 9.807, 10.297, 10.787 m/s^2) and five different sizes (0.0703, 0.07215, 0.074, 0.07585, 0.0777 m) launched at eye-level 30 meters away from the observer in a head-on approach. We chose those values of gravitational acceleration and size using differences of 5% the standard of Earth gravity and 2.5% of the standard size of a baseball to envision reliable values of prior knowledge of gravity and size (however, see Jörges and López-Moliner, 2020b).

Figure 4.4 represents 9 test trajectories exposed to the system (panel A), whereas panels B, C and D represent the range of possible values for each variable available for the observer: retinal size (θ), the elevation angle (γ) and the rate of change of the elevation angle ($\dot{\gamma}$) respectively.

To reproduce the encoding process of a sensory stimulus, we simulated a set of tuning curves covering the range of possible stimulus strengths (stimulus values) for each optic variable 200 ms. after motion onset. In Figure 4.5, the reader can see an illustration of the tuning curves (black curves), the stimulus strength presented to the system (blue vertical line), and an example of the average response by a detector (red curve) for the average TTC under standard conditions of gravity and size (light blue trajectory in Figure 4.4). The detectors simulated for retinal size (θ) covered a range from 0.0024 to 0.0031 rad (SD = 0.00014; representing a 5% WF). The stimulus range covered for the elevation angle (γ) and its rate of change ($\dot{\gamma}$) was from 0.045 to 0.085 rad (SD = 0.00325 rad; 5% WF) and from 0.2 to 0.45 rad/s (SD = 0.01625 rad/s; 5% WF) respectively.

After the detectors were exposed to the stimulus, we obtained the average response probability (r) (solid red line in Figure 4.5) for corresponding neural detectors (Dayan & Abbott, 2001) as:

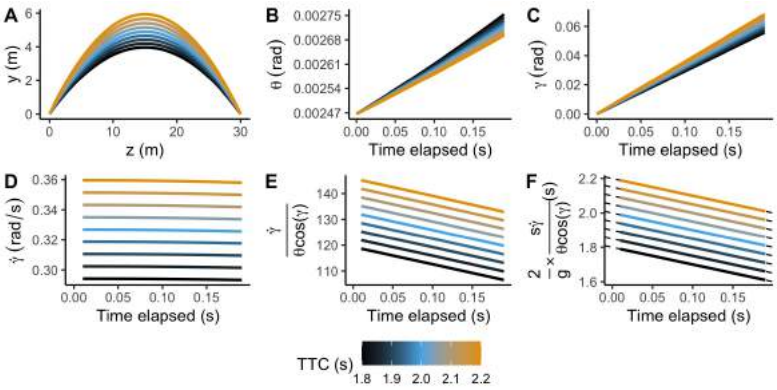

Figure 4.4. Test parabolas used in the simulation. A) Ball´s vertical versus depth position. Panels B, C and D indicate retinal size (θ), elevation angle (γ) and rate of change of the elevation angle ($\dot{\gamma}$) as a function of time for the first 200 ms. of the trajectory. Panels E and F depict the output of the GS model using visual information only and combined with prior information, respectively.

$$p(r) = f(\theta, \gamma, \dot{\gamma})^r \frac{e^{-f(\theta,\gamma,\dot{\gamma})}}{r!} \tag{4.4}$$

In this expression, (f) indicates the mean activity per detector, whereas (r!) corresponds to the factorial response for each detector. In our simulation, the resulting activation varies on each iteration by adding Poisson noise, representing random variability in neural activation.

We simulated 1000 trials per TTC in which the size was the standard of a baseball (0.074 m), and gravity was the standard on Earth (9.807 m/s^2). Once the optic variables were encoded by the detectors simulated, we recovered the most likely stimulus strength presented to the system for each optic variable using a Maximum Likelihood estimate (MLE) procedure. Then, we obtained a value corresponding to the GS model based only on the optic variables retrieved by the encoding procedure.

We compared that output to the GS model's ideal (noiseless) output based only on optic variables for all the simulated trajectories. We obtained two possible sets of decoding responses to select all the potential trajectories that matched the model's output. For the first set of responses, we only used sensory information without resorting to size or gravity priors to decode the

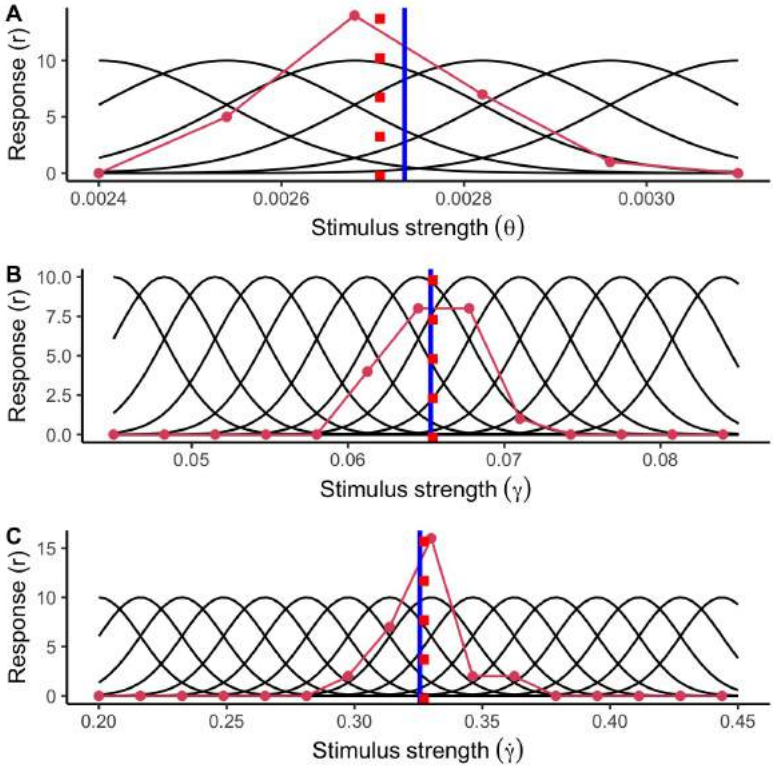

Figure 4.5. Simulated tuning curves of neurons specialized for different values of A) retinal size (θ), B) elevation angle (γ) and C) rate of change of the elevation angle ($\dot{\gamma}$). The blue vertical line indicates the true stimulus strength exposed to the system. The stimulus strength was selected from the standard condition 200 ms. after motion onset (see main text). The red curve indicates the average activation per neuron in a single trial (Poisson noise added). The red dashed lines indicate the stimulus strength inferred by the encoding process.

correct TTC; that is, we used a Maximum Likelihood Estimation procedure. We assumed the correct size and gravity priors to decode the correct TTC for the second set of responses. Then, we selected all potential trajectories that would fall into a relatively low error margin of $\pm$ 5%.

Figure 4.6 depicts the average accuracy per procedure. Here, the performance of a system using only sensory information (MLE; red dashed line) is slightly over the chance level (blue dashed line). On the contrary, using the correct priors (central dots in panels A and B of Figure 4.6) improves the proportion of correctly estimated TTC's substantially. Note that since the relative difference between the different simulated gravity values is higher than those of size, the procedure benefits more from an accurate representation of internalized knowledge of gravity. This example highlights that, despite the inherent ambiguity of sensory information in the optic flow, the use of prior information is a powerful calibration tool to interpret otherwise ambiguous visual information.

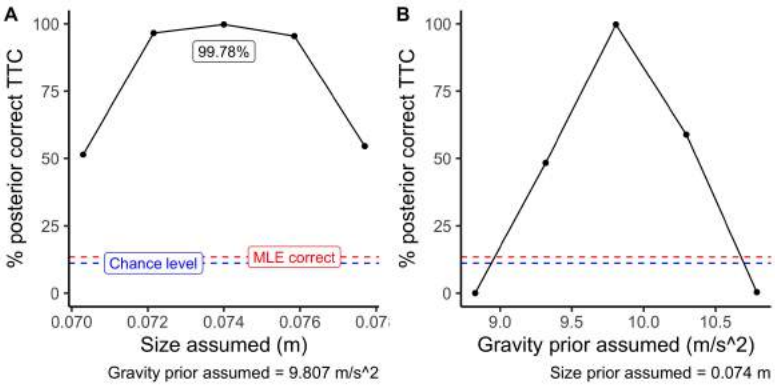

Figure 4.6. Average accuracy per procedure. The blue dashed line indicates chance level (11%), red dashed line indicates the performance of an MLE procedure. The black points indicate the performance assuming different size priors in panel A (g = 9.807 m/s^2 assumed) and assuming different gravity priors in panel B (baseball size assumed).

Accuracy and precision of using gravity and known size

So far, we have shown that calibrating visual information in the light of prior knowledge allows us to draw accurate predictions using the GS model. However, it could still be the case that even if the output is accurate, the visual information in the optic flow is so noisy that an estimation of TTC might not be available.

To evaluate an observer's ability to estimate TTC accurately and precisely from the GS model, we simulated a series of typical spatio-temporal parameters for parabolic trajectories. We simulated parabolic trajectories launching at eye-level at five initial distances (Z_{Init} = 15,20,30,40,50 m.), one contact time (TTC = 2 s; Δ_t = 0.01 s.) and a single radius corresponding to a baseball (0.037 m.). In each case, the endpoint is the origin, that is, the position of the simulated observer.

To evaluate the precision of the output, we introduced independent Gaussian noise to θ, γ and $\dot{\gamma}$ according to their respective Weber fractions (identified with the letter k).

$$\theta_\chi = \theta + N_{(\mu=0;\sigma=\theta\cdot\theta_k)} \qquad \theta_k = 0.05 \tag{4.5}$$

$$\gamma_\chi = \gamma + N_{(\mu=0;\sigma=\gamma\cdot\gamma_k)} \qquad \gamma_k = 0.05 \tag{4.6}$$

$$\dot{\gamma}_\chi = \dot{\gamma} + N_{(\mu=0;\ \sigma=\dot{\gamma}\cdot\dot{\gamma}_k)} \qquad \dot{\gamma}_k = 0.05 \tag{4.7}$$

To test the noise-suppression performance for the GS model, we ran 10,000 simulations for each condition using Equations 4.5 - 4.7. Then, we obtained a Weber fraction timewise as the ratio between the standard deviation of the signal and the mean predicted TTC. Note that the χ version of a variable denotes its noisified version.

Figure 4.7A depicts the output of the GS model for each initial distance (colour code), whereas the inset indicates the predicted temporal error for the ideal (noise-free) output of the model. The GS model provides very accurate estimates for which the maximum error is lower than 10 ms. On its part, Figure 4.7B represents the Weber fraction estimated timewise for the GS model. As a comparison, the Weber fraction for Tau was envisioned

as constant at 10% as reviewed above. The GS model presents a precise output during most of the trajectory (Weber Fraction is always lower than 10%), resulting in an accurate and robust solution to the estimation of TTC in parabolic trajectories comparable to previous Weber Fractions found in the literature (Jörges et al., 2021; Moscatelli & Lacquaniti, 2011).

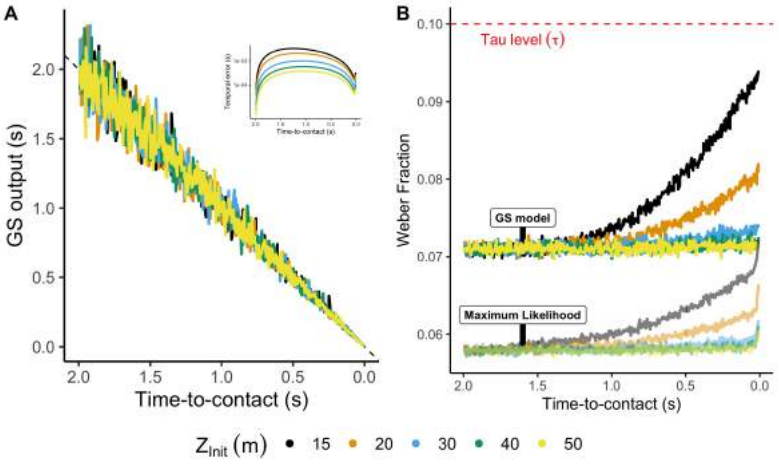

Figure 4.7. A) Noisified estimates of TTC using the GS model for different trajectories. The inset represents the temporal error for the noiseless output of the GS model. B) Weber fraction computed as the ratio between standard deviation and mean of the GS model each frame. The red dashed line indicates the mean Weber fraction of Tau (see main text). The translucid output indicates the Weber fraction of a combination of the GS model and Tau using an MLE procedure.

It is essential to mention here that just as others have already described in the literature, the sources of information to estimate TTC may vary depending on the segments of an approach visible (DeLucia, 2004; DeLucia et al., 2016; Gómez & López-Moliner, 2013). While an initial temporal estimate would be available using the GS model, final interceptive actions would take advantage of more straightforward strategies such as a distance criterion (Gómez & López-Moliner, 2013; López-Moliner & Keil, 2012; Wann, 1996), Tau (Lee, 1976; Shaffer & McBeath, 2005; Zago et al., 2004) or correlates of binocular disparity (Rushton & Wann, 1999).

Following this reasoning, de la Malla and López-Moliner, 2015 partially validated the use of the GS model, showing that early estimates of TTC

based on the GS model could be integrated with the latest estimates derived from correlates of the rate of expansion resulting in an accurate and precise timing mechanism. Mimicking that context, we combined the predictions of Tau and the GS model using a maximum likelihood process (Ernst & Banks, 2002). The output results in a robust solution against sensory noise for the estimation of TTC and timing interceptive actions (see translucid lines in Figure 4.7B).

4.4.2 Generalization of the GS model

The formalization of the GS model assumes that the ball moves in a collision course with the observer (Gómez & López-Moliner, 2013). However, this is not usually the case. Commonly an observer must move to intercept the ball. Therefore, the following question is to what extent the output of the GS model deviates from perfect accuracy for trajectories ending in locations other than the observer's position?

To investigate those cases, we estimated the output of the GS model for trajectories ending at different interception locations. We simulated one initial distance (Z_{Init}= 50 m.) and eight interception points around the observer (see Figure 4.8A). Initially, the GS model provides accurate estimates of the TTC regardless of the position of the observer. Then, in contrast with trajectories on a collision course, the simulation reflects systematic errors in TTC estimation shortly after motion onset if the observer remains stationary (see Figure 4.8B). If the ball falls behind the observer, the rate of change of the predicted TTC decreases. Thus, the model's output overestimates the remaining TTC and vice versa for balls falling ahead (see Figure 4.8B), pointing out that the errors depend on the interception location. In this context, a navigational strategy predicting where and when the ball would be within reach would initially guide the observer towards the wrong position.

Nevertheless, the simulations described so far result in predictions of TTC for an unlikely situation in which the observer is not at the interception location and remains stationary. Usually, an observer would control the ball moving towards the interception area. As a result, an observer's movement would prompt changes in the optic flow. As we will show, simulating the observer's movement, we found interesting properties in the output of the

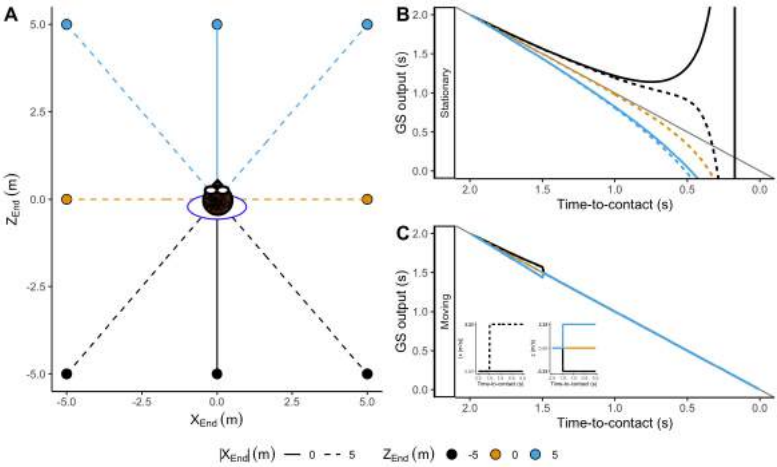

Figure 4.8. A) Ending positions for simulated trajectories around the observer. The lines represent the trajectories followed by the moving observer. Figures B) and C) depict the output of the GS model for a stationary and a moving observer. The line code indicates lateral ending position (X_{End} = 0, 5 (m)). The colour code indicates the ending position in depth (Z_{End} = -5, 0, 5 (m)) . Note that the GS model predicts an underestimation of TTC for balls falling ahead and an overestimation of TTC for those falling behind the observer.

GS model that might indicate the availability of a navigational strategy.

To perform the following simulations, we replicated the previous trajectories. However, in this case, the observer started moving towards the interception location at a constant speed 500 milliseconds after the ball's launch. We chose this moment because catchers generally start running in the right direction 500 milliseconds after the ball's flight has started (Brouwer et al., 2006; McLeod & Dienes, 1993, 1996; Michaels & Oudejans, 1992). The displacement speed of the observer was computed to reach the interception point just in time to catch the ball (see inset in Figure 4.8). Although this pattern of displacement and speed does not correspond precisely to the found in real life (McLeod & Dienes, 1993; McLeod et al., 2006), it is essential to point out that it will be useful for an illustrative purpose.

In Figure 4.8B, the reader can see that when the observer remains stationary for some time in a position other than the interception location, the rate of change of the predicted TTC changes (see Figure 4.8B). When the rate of change in TTC decreases, the ball will fall behind the observer and vice versa. This would signal the need to move and the correct direction in depth. Therefore, departures from the initial rate of change in TTC (slopes different from -1) could be used as a navigational strategy indicating if the observer must move forward or backwards. Then, an observer's movement provides the necessary changes in the optic flow to linearize the predictions of the remaining TTC (see Figure 4.8C). Thus, keeping the prediction linear will ensure that the observer would end up at the interception position in time.

In sum, the above simulations indicate that the model's output is accurate when the observer moves in the correct direction and speed providing the basis for a mechanism to navigate towards the interception location. However, these simulations were performed in a context in which the ball is only affected by the gravitational acceleration. Would a simulation of trajectories under air drag provide equally accurate temporal estimates?

Dynamic effects: Air drag

In real life, the ball is affected by external forces other than gravity, such as air drag, Magnus force or wind currents. These forces deviate the trajectory

from a perfect parabola compared to motion in a vacuum in astonishing ways (McBeath et al., 2008). For instance, previous works indicate that air drag can reduce flight time and distance travelled by a flying ball up to 50% (Adair, 2002; Brancazio, 1985). Therefore, trajectories initially on a collision course with the observer are no longer so after a short period. This pattern would potentially preclude the use of different algorithms for estimating the TTC, such as the Tau or GS model.

It has been argued that the self-regulatory nature of information-based strategies can efficiently deal with dynamic effects in a parabolic trajectory, provided that continuous visual information is available. In contrast, it is commonly argued that an internal model assuming a constant gravitational acceleration would be insufficient to account for dynamic forces such as air drag (Fink et al., 2009). To account for air drag, an internal model would have to gain access to a drag coefficient, mass and size for every single object and environment dynamically, which limits a massive application (Craig et al., 2006). Furthermore, it seems at odds with the fact that most people think that objects fall at the same rate despite their mass or volume (Oberle et al., 2005). However, explicit knowledge of physics may not affect performance in action-related tasks (Flavell, 2014; Reed et al., 2010). Following this reasoning, in our view, predictions using priors would only include variables facilitating the interpretation of the most generic case of natural law or parameters for a given task. In the following, we will show how the GS model, which relies only on gravity and size priors, can predict the remaining TTC reliably for the general case of trajectories under gravity and air drag conditions.

Unlike the gravitational force, which exerts the same force for different projectiles, air resistance depends on several factors: ρ, the viscosity of the environment surrounding the object; C_d, a drag coefficient relative to the texture and shape of the projectile essentially; r object's radius and v, the tangential speed of the object estimated dynamically. To simulate the effects of air drag on a parabolic trajectory, we followed the procedure described in Timmerman and van der Weele (1999) and Gómez and López-Moliner (2013). We simulated different trajectories under two different conditions: *gravity only* and *gravity + air drag*. Air viscosity around the ball (ρ) was set to 1.225 kg/m^3 (value at sea level), and C_d was set to 0.346 or 0.4

for baseball and soccer balls, respectively (Alam et al., 2010; Kagan & Nathan, 2014). We introduced one initial vertical speed $Vy_0 = 9.807m/s$ (2 seconds of flight time under gravity only conditions) for each trajectory and corresponding approaching speeds (Vy_0 = 7.5,15,25 m/s) different balls launched at the origin. In this simulation, we did not include horizontal displacements.

In Figure 4.9A, the reader can see how air drag influences the trajectory described by the ball in both: the spatial and the temporal domain. In Figure 4.9A, the grey trajectory represents the trajectory followed by a ball under gravity-only conditions, whereas the blue and red trajectories indicate the trajectories followed by baseball and soccer balls including air drag in the simulation. For soccer balls, the effect of air drag is more pronounced mainly due to a larger cross-sectional area against the air. Note that since the initial vertical speed was the same for all the trajectories, the differences in flight time and distance travelled can be attributed to the different approaching speeds.

Thus, how well can the GS model estimate the remaining TTC in trajectories, including air drag? To answer this question, we simulated the output of the GS model for the worst-case scenario previously simulated. In that case, the GS model will yield the least accurate predictions. As shown in Figure 4.9A, the trajectory most affected by air drag is when a Soccer ball moves at the highest horizontal speed (Vz_0 = 25 m/s).

To test the model's accuracy, we used three different situations. In the first one, the observer is stationary at the interception position under gravity + air drag conditions (black dot in Figure 4.9A). In the second, the observer is stationary at a midpoint between the fall point under gravity-only and gravity + air drag conditions (black dot in Figure 4.9B). Finally, we simulated a situation in which the observer is at the same "mid-point". However, in this case, the observer moves towards the intercept point at a constant speed (8.13 m/s), 500 ms after motion onset (green arrow in Figure 4.9A).

The output of the model for corresponding situations is depicted in Figure 4.9B-D. In all cases, the GS model reflects initial temporal errors corresponding to the difference in flight time between trajectories under gravity-only and gravity + air drag conditions (see annotations within

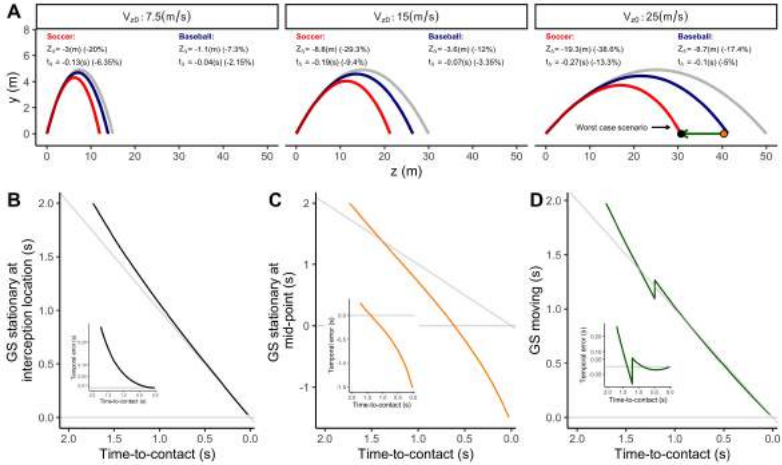

Figure 4.9. A) Lateral view of different parabolic trajectories under gravity (grey lines) and gravity + air drag conditions for two different balls (red: soccer ball; blue: baseballs) and three initial approaching speeds (different panels). The figure annotates the difference in distance travelled (Z_Δ) and flight duration (t_Δ) compared to a trajectory only considering gravity. The black and orange dots in the third panel indicate the position of the corresponding simulated observer in panels B or C and D, respectively. The green arrow indicates the displacement simulated in panel D. Panels B), C) and D) indicate the predicted TTC using the GS model for different simulated observers in the worst-case scenario simulated. Insets depict the corresponding temporal errors using the predictions of the GS model.

Figure 4.9A). When the observer stands still in the interception location (see Figure 4.9B), the GS model presents a high degree of accuracy during most of the trajectory. For instance, 0.5 seconds before the collision, the output converges to temporal errors of about 10 to 20 milliseconds. However, if the observer stands still at the midpoint (see Figure 4.9C), the model's output deviates severely. In this case, the rate of change in the predicted TTC remains consistently lower than -1. In principle, this pattern could inform the observer that the interception point would be ahead of their position. In contrast, when the observer heads towards the interception location (Figure 4.9D), the model yield accurate predictions.

These simulations provide evidence that the output of the information included within the GS model provides accurate and actionable predictions of the remaining TTC when the observer remains stationary in the interception location or displaces towards the interception location. Therefore, it could be used as a navigational strategy or to plan the final interceptive action even when air drag is present.

Benefits and limitations of the generalization of the GS model

The GS model, like Tau, provides temporal information that may involve certain predictive benefits compared to the error-nulling strategies within the outfielder problem. Nevertheless, it also has some limitations that will be addressed in this section.

First, the GS model is much more robust to sensory noise than Tau (Gómez & López-Moliner, 2013). It uses the rate of change of the elevation angle ($\dot{\gamma}$) to estimate the TTC instead of a much noisier variable, the rate of expansion ($\dot{\theta}$) upon which Tau relies.

Second, Tau only provides accurate estimates when the ball is moving at constant speed towards the observer. In contrast, the GS provides accurate estimates at launch independently of the observer's position, which would provide an initial accurate temporal information useful for planning the action.

Third, the GS model overestimates the TTC for trajectories under air drag as a function of the difference in flight time between trajectories under gravity-only and gravity + air drag conditions. However, our simulations show that in combination with the observer movement, temporal errors

of less than 50 milliseconds are possible one second before collision (see Figure 4.9D).

Fourth, in contrast with previous error-nulling strategies, the GS model provides temporal information and can help compensate for the temporal delays and occlusions due to its predictive nature. The TTC can be used to adjust locomotion speed, inform of the remaining TTC under restricted visibility conditions or plan the final manual interception. On its part, the rate of change of TTC could be used to adjust the direction of movement. In this sense, both signals are complementary. However, the viability of the latter requires being able to detect changes in the rate of change of the predicted TTC, which needs further research. Furthermore, the detection of variations in the rate of change of TTC will likely incur delays. Therefore, future studies should study to which extent the simultaneous TTC signal can compensate for these delays.

Finally, the prediction from the GS model shares some limitations with Tau; that is, dealing with non-spherical objects such as Rugby balls or Frisbees would need further elaboration. However, some studies found that those errors can be cancelled out by adding binocular information at the latter stages of catching (Gray & Regan, 1998).

4.5 Evidence of prediction in eye behaviour and manual interception

The main problem to find support for model-based controlled behaviour is that, when possible, the observer would keep track of the trajectory continuously (Oudejans et al., 1999; Postma et al., 2014). Indeed, this is the case of the outfielder problem, for which there is only anecdotal evidence of successful catchers directing away the gaze from the ball (Chodosh et al., 1995). In this context, accurate actions would not allow us to discriminate between information-based and model-based control directly (Belousov et al., 2016). Because of that, we need to scrutinize scenarios in which simple solutions such as heuristics or mappings between sensory information and temporal correlates for temporal estimation or action initiation are not available (Zhao & Warren, 2015).

One possibility to unveil the need for prediction in action control is to

manipulate the target's visibility. It is the most widely used experimental manipulation to study the predictive nature of behaviour in interception (Brenner & Smeets, 2011; de la Malla & López-Moliner, 2015; López-Moliner et al., 2010; Sharp & Whiting, 1974; Spering et al., 2011; Whiting & Sharp, 1974). Nevertheless, more natural conditions are essential to understand how an observer could use temporal estimates to guide their action. In natural conditions, our gaze is often shifted to different locations to gather information that may be relevant shortly (Hayhoe et al., 2005). In other contexts, a player would divert her gaze to check for deviations caused by balls' bouncing (Diaz et al., 2013) or confirm whether the ball was appropriately hit (Mann et al., 2013). Some studies in manual interception and temporal estimations gave the observer complete freedom to decide which part of a trajectory they wanted to exploit visually while dealing with alternative tasks (Aguado & López-Moliner, 2021b; Faisal & Wolpert, 2009; López-Moliner & Brenner, 2016). In those cases, where and when the observer averts the gaze from the target may provide valuable clues about the most relevant pieces of information according to task demands. Therefore, future studies might investigate when people prefer to divert the gaze from the ball while moving towards the interception location.

In some cases, it has been suggested the existence of privileged portions of the trajectory available for an observer to judge TTC. For example, in juggling or catching a ball, looking at the apex would provide the most relevant information (Todd, 1981; J. Watson et al., 1992; Whiting, 1968). However, a closer look at experimental data indicates that an observer does not actively search for a particular position in the parabola. Instead, prefers to use fixed temporal viewing windows generating priors during the task. These priors could then be used to weight visual information or estimate TTC when sensory information is unavailable (Aguado & López-Moliner, 2021b; Amazeen et al., 1999; López-Moliner & Keil, 2012).

For example, most studies show acceptable catching performance in manual interception tasks for short flight durations no matter the section of the trajectory viewed. However, visual information had to be captured at least 200 ms before the catch to avoid sensorimotor delays (López-Moliner & Brenner, 2016; López-Moliner et al., 2010; Sharp & Whiting, 1974). For longer flight durations (up to 2 seconds), catching performance describes

an inverted U shape with respect to flight duration (Amazeen et al., 1999; Sharp & Whiting, 1974). If the observer can only see the ball well in advance, performance would be low because the predictions decay rapidly (Aguado & López-Moliner, 2021b; Binsted et al., 2006).

Zhao and Warren, 2015 reasoned that in the case of short flight durations, part of an observer's performance could be explained by the observer having learnt some of the regularities of a predictable trajectory mapping optic variables with a temporal correlate. Still, this would indicate the usefulness of developing priors during the task, which would be exploited when online visual information is not available. However, the fact that an observer exploits visual information when optic mappings are available indicates that they prefer to update their predictions based on the latest available visual information and combine it with evidence from previous knowledge (Binaee et al., 2016; Mazyn et al., 2007). In the end, having a rough prediction is better than none (Brenner & Smeets, 2018).

In this line, de la Malla and López-Moliner, 2015 proved that general rules of integration apply to the estimation of the TTC, which means: the observer integrates past and concurrent information to optimize the precision of temporal responses in a continuous fashion (Dimitriou et al., 2013; Liu & Todorov, 2007; Todorov, 2004). Assuming this is true, we can use Kalman filters to predict an observer's estimation of TTC and response variability. A Kalman filter (Kalman, 1960) is a Bayesian tool that estimates the state of a system combining new noisy estimates, a prediction from prior measurements and a prior knowledge of how the system behaves. Using this technique, we could estimate both the accuracy and precision of online measurements for temporal judgments, manual interceptive tasks, and more general interceptive tasks such as the locomotion within the outfielder problem.

4.6 Future research

One of the main objectives of this work is to highlight the potential role of prior knowledge in calibrating visual information in terms of actionable predictions such as TTC. In our view, drawing predictions based on prior knowledge is not just a reliable and accurate way to predict future states of

the environment but also helps us override the need to use unreliable optic cues (Cutting & Vishton, 1995). For instance: expansion rate ($\dot{\theta}$) and thus Tau (τ) might not be available at large distances, optic acceleration ($\ddot{\gamma}$) is hardly discriminable by humans and would not allow for the continuous control of action (Werkhoven et al., 1992) even though some studies claim that the values of optical acceleration available are large enough to be detected (Babler & Dannemiller, 1993; Zaal et al., 2012); lastly, the literature is mixed with regards to the benefits of providing binocular disparity. Because of these reasons, here we advocate for using the rate of change in the elevation angle ($\dot{\gamma}$), which is very precise, in combination with an internalized knowledge of gravity and physical size for the estimation of TTC in the GS model.

Here, the GS model provides different contexts to test the information included. For example, within the GS model, each contextual piece of information, gravity or size, is either in the denominator or the numerator. Hence, introducing proportional changes in the parameters governing the trajectory would result in proportional errors in the estimates of the remaining TTC. In a similar line, Jörges and López-Moliner, 2017 showed that an observer might be able to extract information about the approaching speed of a ball through estimations of the rate of change of the elevation angle prior knowledge of gravity. Therefore, different values of gravity governing a trajectory should influence the prediction of the interception location.

Moreover, using TTC discrimination tasks, it could be possible to study if an observer can detect differences between trajectories under gravity-only conditions and gravity + air drag conditions. Our simulations indicated a Weber Fraction of about 7% for the GS output (see Figure 4.7). Therefore, the difference in TTC should be above the discrimination threshold in some cases, as depicted in Figure 4.9. Furthermore, decision tasks based on an observer's ability to decide if there is enough time to perform alternate tasks (e.g. looking for teammates or running towards the interception area) from early visual information might be essential to test the availability of temporal estimates as a parameter to plan action for a broader range of interceptive actions.

On another note, it might be interesting to investigate the use of the GS model as a navigational mechanism. Since the GS model does not specify

the interception location to plan interception in advance, we discovered a continuous coupling to keep a constant rate of change of the predicted TTC. To test if an observer would adapt locomotion to a constant rate of change, we should introduce players into contexts in which the value of gravitational acceleration or ball size do not correspond with the parameters assumed a priori. As introduced above, changes in the parameters would result in estimation errors of the remaining TTC. Thus, these manipulations would lead to predictions of the path followed by the observer. Nevertheless, to be able to use such a strategy, an observer might be able to detect deviations from different rates of change in TTC. To our knowledge, there is no previous work providing figures about how well people detect changes in TTC. Thus, our ability to detect differences in the rate of change and the time required to do so will need to be studied in future works.

To generate the suggested experiments, we need immersive and realistic spaces. Virtual scenarios will provide ecologically valid contexts to evaluate to what extent predictions influence interception. To do so, the use of wireless head-mounted displays (HMD) and portable eye-trackers will be essential. Head-mounted displays insert the participants into rich and controlled environments already being used to train professional sports players (Gray, 2017; D. Harris et al., 2020; Zaal & Bootsma, 2011). Combining this technique with built-in eye-tracking systems provides access to how players interact with the environment to gather relevant visual information (Binaee et al., 2016; Moran et al., 2018). Those findings would still need to be replicated in real life under full-cue conditions. However, augmented reality devices are becoming more and more accessible and are likely to become more widespread. Those results may not be fully transferable to real life. However, it still would provide us information about human performance interacting with increasingly in-demand devices with potential applicability in a growing industry, eSports.

5 Study 3: Flexible viewing time when estimating time-to-contact in 3D parabolic trajectories

Obtaining reliable estimates of the time-to-contact (TTC) in a 3D parabolic trajectory is still an open issue. A direct analysis of the optic flow cannot make accurate predictions for gravitationally accelerated objects. Alternatively, resorting to prior knowledge of gravity and size can provide accurate estimates of TTC in parabolic head-on trajectories, but its generalization depends on the specific geometry of the trajectory and particular moments. The aim of this work is to explore the preferred viewing windows to estimate TTC and how the available visual information affects these estimations. We designed a task in which participants, wearing an HMD had to time the moment a ball in a parabolic path returned at eye-level. We used five trajectories for which accurate temporal predictions were available at different points of flight time. Our results show that our observers can predict both the trajectory of the ball and TTC based on the available visual information and previous experience with the task. However, the times at which our observers chose to gather the visual evidence did not match those in which visual information provided accurate TTC. Instead, they looked at the ball at relatively fixed temporal windows depending on the trajectory but not of TTC.

This study has been published as: Aguado, B., & López-Moliner, J. (2021b). Flexible viewing time when estimating time-to-contact in 3D parabolic trajectories. *Journal of Vision*, *21*(4), 9. https://doi.org/10.1167/jov.21.4.9

5.1 Introduction

The time remaining before an object reaches a point of interest is called TTC (time-to-contact). To estimate this parameter accurately is of great importance, as it can be used for multiple actions such as avoiding collisions, intercepting moving targets and, more generally, regulating one's own speed.

Previous literature has focused on the study of TTC for objects under different visual conditions such as objects moving at a constant speed in the fronto-lateral plane (Bootsma & Oudejans, 1993; Tresilian, 1994) or moving towards the observer (Heuer, 1993; Lee, 1976; Wann, 1996). In contrast, during the last decades, others have focused on the study of gravitationally accelerated objects in free-fall (Lacquaniti & Maioli, 1989; McIntyre et al., 2003; McIntyre et al., 2001; Zago et al., 2004), parabolic motion in head-on trajectories (de la Malla & López-Moliner, 2015) or fronto-parallel ones (Joerges et al., 2018; Jörges & López-Moliner, 2019). However, the estimation of TTC for objects describing parabolas in the more general case has not been systematically addressed. This is probably due to the complex mapping between the distal three-dimensional trajectory and the projected optic variables. The same optic pattern can be caused by a multitude of sources in the external world (Pizlo, 2001) rendering the interpretation of the real source of stimulation an ill posed problem known as the inverse problem of vision (Kersten et al., 2004). This problem would compromise an agent's performance based on optic information alone.

A paradigmatic case in which the computation of TTC for a 3D flying object is key is the so-called *outfielder problem*. In baseball, players known as *outfielders* must catch a flying ball at a specific time and location avoiding ground contact. The distances involved in this task and the size of the ball render binocular cues and retinal expansion non-discriminable. Therefore, it is usually assumed that only a reduced set of monocular cues are available to guide action (Cutting & Vishton, 1995; Wilson et al., 2013; Zago et al., 2008). To control the ball, some people advocate to keep task-relevant sources of visual information invariant (Chapman, 1968; Fink et al., 2009; McBeath et al., 1995). For example, Chapman (1968) realized that all an observer had to do to control the ball is to keep the elevation angle (γ), that

is, the vertical angle between ball's position and an observer's eye-level at an increasing rate without passing over an observer's head. These strategies override the need to estimate parameters of the task such as the interception area or TTC (Shaffer & McBeath, 2005). Nevertheless, this solution is strongly dependent on concurrent sensory feedback. Because of that, it fails to explain catching behaviour when the observer must turn the gaze from the ball to run directly towards the interception area or look around for a partner (Belousov et al., 2016).

It is known that continuous visibility is not particularly necessary to make a catch (Chodosh et al., 1995). Instead, all that seems necessary to catch a ball at hand is to see a portion of the trajectory large enough to integrate positional and motion estimates (Elliott et al., 1994) before the last 200 milliseconds in order to avoid sensorimotor delays (Keele & Posner, 1968; López-Moliner & Brenner, 2016; López-Moliner et al., 2010; López-Moliner & Keil, 2012; Sharp & Whiting, 1974). In this line, an alternative question arises: is there any moment when viewing the ball is most beneficial?

Previous mathematical analysis of the information provided in the optic flow suggested that there are certain privileged positions that an observer can exploit to judge the remaining TTC. For example, in the case of juggling or catching a ball in a parabolic trajectory, kinematic information around the apex would provide privileged information to predict the remaining TTC (Todd, 1981; J. Watson et al., 1992; Whiting, 1968). This hypothesis was tested showing that, actually, the observers do not actively search for a *particular position* during the course of the parabola. Instead, they use fixed visibility windows to time the interceptive action when online visual information is not available or reliable (Amazeen et al., 1999; López-Moliner & Brenner, 2016; López-Moliner et al., 2010). However, a problem with the studies reported above is that the balls were always thrown towards the observer and the flight time was relatively short (e.g., less than a second). Therefore, actual predictions could be overridden by learnt mappings between visual information and the remaining TTC without the need to invoke computations by an internal model (Zhao & Warren, 2015). In fact, the mathematical analysis mentioned above was carried out in terms of Cartesian variables which, unlike optic variables, are

not readily available to human experience. Therefore, it is first necessary to explain how to move from optic to Cartesian variables in order to estimate the remaining time to contact.

Over the last years, a growing body of literature has advocated for the combined use of optic and prior information for the estimation of TTC. *A priori* knowledge of an object's size (Hosking & Crassini, 2010; López-Moliner et al., 2007; López-Moliner & Keil, 2012) and gravitational acceleration (Brouwer et al., 2006; McIntyre et al., 2001; Saxberg, 1987b) has been suggested to calibrate the optic space into actionable estimates of TTC (Gómez & López-Moliner, 2013) and approaching speed (López-Moliner et al., 2007). Along these lines, Gómez and López-Moliner (2013) proposed the so-called GS model. The GS model resorts to a combination of contextual variables (ball size s and gravitational acceleration g) that are assumed to be constant and known by the observer along with time-dependent monocular cues present in the optic flow. The GS model resorts to retinal size (θ), the elevation angle (γ), and it's rate of change ($\dot{\gamma}$) to obtain accurate estimates of TTC for head-on trajectories.

$$\mathrm{TTC}_{\mathrm{GS}} \approx \frac{2}{g}\frac{s}{\theta}\frac{\dot{\gamma}}{cos(\gamma)} \tag{5.1}$$

A priori known ball size (s) combined with retinal size (θ) maps retinal variables into Cartesian metrics. On its part, gravitational acceleration (g) normalizes the rate of change of the elevation angle ($\dot{\gamma}$) into meaningful estimates of TTC under arbitrary gravity values. As a result, the GS model is an algorithm capable of signalling accurately TTC for trajectories on a head-on approach throughout the entire flight (Gómez & López-Moliner, 2013). This is the case for the trajectory represented by a grey (thinner) line in Figure 5.1.

However, when a projectile is not landing on the observer, the GS model predictions are no longer accurate during all the ball flight, because the elevation angle (γ) does not increase during the whole trajectory. As an example, in Figure 5.1A the reader can see a top-view representation of 5 different parabolic trajectories. Each trajectory presents a different lateral offset and two different flight durations (see inset in Figure 5.1A). Figure 5.1B and Figure 5.1C represent respective predictions and predicted errors of the remaining TTC across time using the GS model. There the

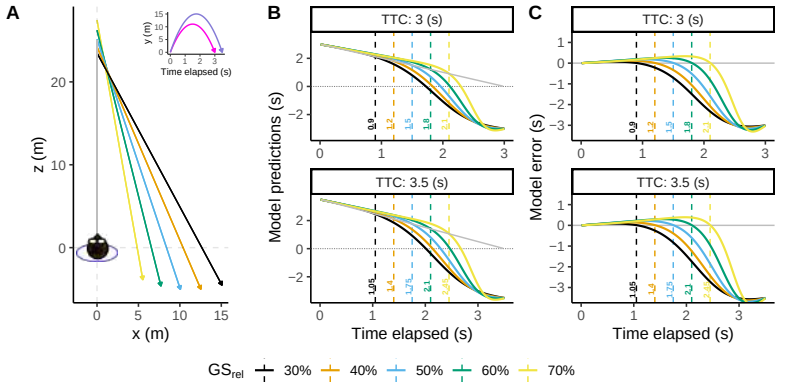

Figure 5.1. A) Top-view representation of the five different trajectories tested in the study and a head-on approach (grey line). B) Predictions of remaining TTC for each trajectory as a function of time elapsed since motion onset using the GS model. Grey line constitutes near-perfect accuracy for a head-on trajectory using the GS model. C) Prediction error for each trajectory using the GS model. Positive errors indicate an overestimation of the remaining TTC and vice versa.

reader can see that the predictions drawn by the GS model depend on the geometry of the corresponding trajectories (GS_{rel}) and flight duration (TTC). When the ball is launched at eye-level, the GS model signals the remaining flight time accurately. Afterwards the remaining flight time is overestimated (positive errors). At a certain time for each trajectory, the predictions become accurate for a short period of time (indicated in Figure 5.1B and Figure 5.1C with dashed lines). After that, the model severely underestimates the remaining flight time rendering the predictions invalid.

Figure 5.1B shows that these model predictions could be used as a starting point to analyse if people can exploit the information used by the model to estimate TTC, which can be performed accurately at different proportions of flight time (see the key legend in Figure 5.1). As an example, the visual information contained in the model would allow to estimate the remaining TTC accurately by 1.5 or 1.75 seconds after launch for the blue trajectory depending on the duration of the flight. These privileged time points correspond with a 50% of total flight time passed (as indicated by $GS_{\text{rel}} = 50\%$). In principle, if an observer is sensitive to the information

included in the GS model and is able to improve the performance based on feedback by using such privileged windows, it seems reasonable to assume that the observer will look actively at the ball during these temporal windows.

In this study we will explore whether observers are able to predict TTC for parabolic trajectories of 3- and 3.5-seconds restricting ball's visibility to an initial window of 300 milliseconds and a second mid-flight window of 400 milliseconds that depends on the observer's actions in each trial. With this manipulation, we will explore the preferred windows for the estimation of the remaining flight time. In that regard, the GS model can be used as a reference since provides a different privileged temporal point to estimate the remaining flight time for each trajectory. Then, we will explore if the observers can estimate the remaining flight time with the visual information available at the moments when they look at the ball. This will allow us to test if their predictions conform to the GS model.

5.2 Methods

5.2.1 Participants

In this experiment, we tested ten participants (n = 12; 4 self-identified women). They were between 21 and 32 years old. They had normal or corrected-to-normal vision. All of them were naïve to experimental goals and volunteered to take part in the experiment. This study is part of an ongoing research program that has been approved by the local ethics committee of the University of Barcelona in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

5.2.2 Materials

Apparatus

The experiment was run by an Intel i7-based PC (Intel, Santa Clara, CA). The stimuli were rendered by an NVIDIA GeForce GTX 1070 and sent for display to an HTC Vive Pro head-mounted display at 90 Hz per eye. HTC Vive Pro has a Field of view (FOV) of about 100 horizontal degrees

and 110 vertical degrees. Head position and rotation was recorded by two SteamVR BASE STATION 2.0 at 90 Hz (see a representation of the setup in Figure 5.2A).

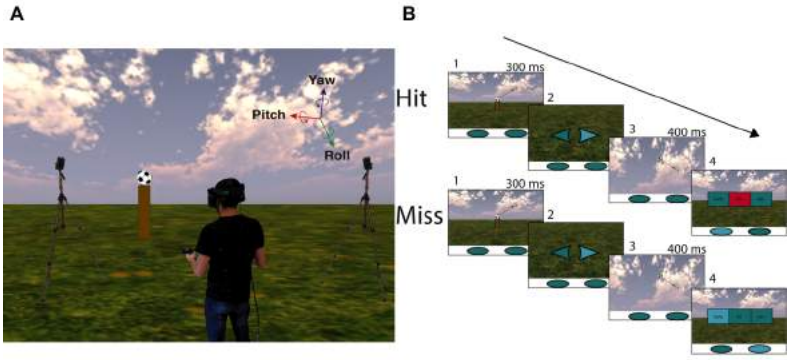

Figure 5.2. A) Representation of the back view inside the environment of the experimental setup. B) Representation of the course of two typical trials. 1: The ball is visible for 300 milliseconds after launch. 2: When looking at the floor, an arrow lights up indicating the correct controller for using in the timing task. 3: The viewer decides when to look up again for the ball (visible for 400 milliseconds). 4: The observer receives feedback about the temporal task ("early," "good" or "late") and the use of the correct controller (blue or red colour panels).

Head rotation was defined using a 3-axis system: yaw, pitch and roll. Assuming that the observer keeps the roll axis constant, the rotation over the pitch axis corresponds to a change in gaze vertical direction and rotation over the yaw axis corresponds to horizontal changes in gaze direction. Since we will be referring to both ball's angular position and gaze direction, hereafter we will refer to yaw as β_{gaze} (horizontal) and pitch as γ_{gaze} (vertical) to ease the interpretation of results (see Figure 5.3). Note that gaze is inferred from the position of the head, no eye-tracking device was used.

Latency between head movement and visual feedback is about 30 milliseconds whereas controller latency is about 15 milliseconds (Chénéchal & Goldman, 2018). Both should render almost negligible effects in the case of the present experiment because we are using flight times of 3 and 3.5 seconds.

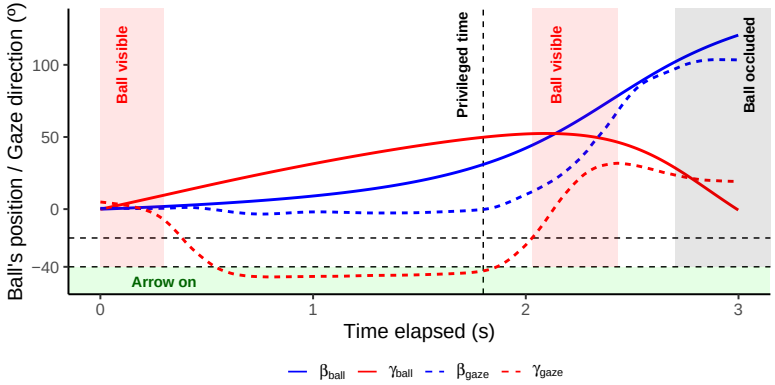

Figure 5.3. A) Representation of ball's position (solid lines) and gaze direction (dashed lines) across time. Blue and red lines indicate horizontal and vertical angles, respectively. The red area indicates ball's visibility windows. Note that the mid-flight visibility window depends on an observer's gaze (threshold at -20 degrees). The green area indicates the visibility of the arrows on the floor (threshold at -40 degrees). The ball was always occluded 300 milliseconds before returning at eye-level again (horizon).

5.2.3 Stimulus

Our stimulus consisted of a Soccer ball (radius = 0.11 m) moving frontally along the 3 dimensions on parabolic trajectories. The ball was frontally and vertically aligned with the observer's eye level at the beginning of each trial to account for posture changes during the experiment.

Flying time was set to 3 seconds for 5 out of 6 trials. However, in order to prevent observers from developing rhythmic responses, a second flight time of 3.5 seconds was randomly interleaved in a proportion of 1 out of 6 trials.

We explored the 10 different trajectories shown in Figure 5.1A corresponding with a combination of 5 different initial positions in depth (Z_{init}= 23.49, 24.23, 25.16, 26.24, 27.47 m.) and two flight times (see inset within Figure 5.1A). Each initial distance corresponds with a lateral final position relative to the observer (X_{end}= 15.09, 12.53, 10.06, 7.71, 5.53 m. either *left* or *right*) and depth position (Z_{end}= -4.53, -5.01, -5.03, -4.63, -3.87 m.) both with a final height at eye level. Each trajectory was tailored to

predict perfect accuracy by the GS model at different temporal moments corresponding to a 30, 40, 50, 60, 70 % of the flight time elapsed, that is, 0.9, 1.2, 1.5, 1.8, 2.1 s. and 1.05, 1.4, 1.75, 2.1, 2.45 s. for trajectories of 3 and 3.5 seconds of flight time respectively.

A detailed scene was stereoscopically displayed providing cues for relative distance, retinal size (θ) and cardinal motion angles (horizontal: β_{ball}, vertical: γ_{ball}). Note that from now on we will use β_{ball} and γ_{ball} to denote ball's angular position in the horizontal and vertical axis with respect to the observer and ball's initial position. See Figure 5.3 for a combined representation of ball's angular position and gaze direction. Figure 5.4 represents β_{ball} and γ_{ball} across time in our experimental trajectories. Note that horizontal angles larger than 90 indicate that the ball is behind the observer under the initial frame of reference assuming that the observer does not rotate. Gravitational acceleration was set at 1g (9.807 m⁄s^2, the standard at sea level). Complex dynamic effects such as air resistance, and Magnus effects were neglected. Therefore, horizontal and depth movement remained constant during the same trial. No embedded rotation was simulated.

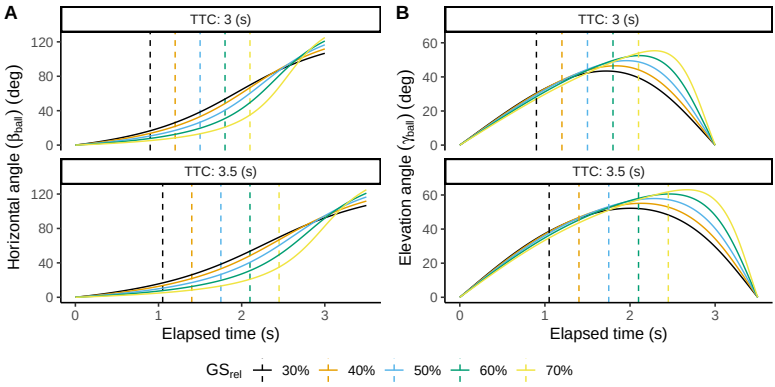

Figure 5.4. Horizontal and vertical ball's angular position for the trajectories present in this study under both flight durations. Dashed vertical lines indicate the privileged time points specified by the GS model.

5.2.4 Procedure

The main events present in the task can be described as (for a graphic representation see Figure 5.2B and Figure 5.3):

1. A Soccer ball appeared frontally aligned with the observer. After 450-550 milliseconds standing still on a pole, the ball was launched and remained visible for 300 milliseconds (first visibility period in Figure 5.3). Note that the ball keeps moving in space even when non-visible.

2. After this 300 milliseconds interval, the participants were instructed to look at the floor, where there will be an arrow lighting up to indicate the proper commander in that trial to perform a temporal judgment. The arrows remain visible if the head's angle is less than -40 deg in γ_{gaze} (γ_{gaze} within the green zone in Figure 5.3).

3. While looking at the floor, they were instructed to freely decide when to look for the flying ball. As soon as they looked up (crossed a fixed threshold at -20 deg. of γ_{gaze}), the ball reappeared and remained visible for a fixed period of 400 milliseconds (second visibility period in Figure 5.3).

4. The ball was always occluded 300 milliseconds prior to returning at eye-level (horizon). After ball's occlusion, the observers had to estimate when the ball would return at eye-level by pressing a button with the proper commander. A trial was considered as a hit if the error of the temporal judgement was within $\pm$ 50 milliseconds (3.3% of 3 seconds). We provided feedback for both, the use of the proper commander and whether the response corresponded to a hit or a miss (early or late response).

All subjects completed 6 blocks of 120 trials in length. Each trajectory was presented 24 times per block (20 test trials and 4 control trials in which flight time was 3.5 seconds). The participants completed between 15-30 training trials before the main experimental procedure in order to familiarise themselves with the task. The researcher did not act as a model for the participant, that is, he did not solve the task in the presence of the

participant. Instead, the participants were always encouraged to "*explore different strategies*" and encouraged to "*freely choose when to look up for the ball to obtain relevant information for the timing task*".

5.2.5 Data Preparation and Analysis

Data preparation and analysis was performed using *R* (R Core Team, 2020). We filtered out outliers using a 1.5 IQR interval based on the response time grouped by subject, trajectory, and flight duration. This procedure removed 342 trials (3.96% of the total).

We checked that ball's relative direction with respect to the observer did not influenced the preferred viewing time ($t(8295.7) = -0.403$, $p = 0.687$) or response time ($t(8263.8) = -0.459$, $p = 0.646$), the two main variables in our study. Then, for the sake of simplicity, we used the absolute value of all the spatial related quantities for further analysis and visual representations.

Then, we tested our hypothesis using Linear Mixed Modelling functions from *lme4* (Bates et al., 2015) package. We used Linear Mixed Models to discriminate between effects in the whole population (*random-effects*) and effects for each experimental condition or subgroup in the population (*fixed effects*). Furthermore, we used a Deming regression present in the *mcr* (Manuilova et al., 2021) R package to fit linear models with measure errors in both x- and y- axis.

5.3 Results

5.3.1 Could the observers predict the trajectory?

In our experiment, the ball reappeared in the second period of visibility for just 400 milliseconds. Therefore, the observers should predict future ball's position picking up information within the first 300 milliseconds.

Due to our experimental design, the ball reappeared when vertical gaze direction (γ_{gaze}) crossed a fixed threshold of -20 deg. Thus, we cannot use this variable directly to check the quality of the prediction. Alternatively, if our participants were able to predict ball's vertical position, the vertical rate of change in gaze ($\dot{\gamma}_{\text{gaze}}$) at ball's reappearance would be linearly related to the position of the ball in the vertical axis (γ_{ball}). To test this hypothesis, we

adjusted a Linear Mixed Model with ball's vertical angular position (γ_{ball}) as a predictor (fixed effect) of the vertical rate of change in gaze ($\dot{\gamma}_{\text{gaze}}$).

Furthermore, if our observers were able to predict ball's position in the horizontal axis, the horizontal position of the ball (β_{ball}) at reappearance must be linearly related to the horizontal direction of gaze (β_{gaze}). To test this hypothesis, we adjusted a Linear Mixed Model in which we used ball's horizontal angular position (β_{ball}) as a predictor (fixed effect) of horizontal gaze direction (β_{gaze}).

For both models we introduced participant as random effects. The Linear Mixed Models were specified as:

$$\beta_{\text{gaze}} \sim \beta_{\text{ball}} + (1|\text{Participant}) \tag{5.2}$$

$$\dot{\gamma}_{\text{gaze}} \sim \gamma_{\text{ball}} + (1|\text{Participant}) \tag{5.3}$$

To check whether our participants used their predictions to guide their gaze, we compared the previous Test models with Null models only including the random term:

$$\beta_{\text{gaze}} \sim (1|\text{Participant}) \tag{5.4}$$

$$\dot{\gamma}_{\text{gaze}} \sim (1|\text{Participant}) \tag{5.5}$$

A Likelihood Ratio Test showed that both Test models are significantly better at predicting gaze direction for the horizontal meridian ($\chi^2(1) = 1460.82$, $p < 0.001$, $BIC_{\text{Null}} = 62443$, $BIC_{\text{Test}} = 60991$) and the vertical meridian ($\chi^2(1) = 12.70$, $p < 0.001$, $BIC_{\text{Null}} = 89759$, $BIC_{\text{Test}} = 89756$) indicating that our participants were able to predict different ball positions across trajectories. This pattern can be interpreted in Figure 5.5, in which, horizontal gaze direction (β_{gaze}) and vertical rate of change in gaze direction ($\dot{\gamma}_{\text{gaze}}$) are linearly related to the position of the ball in both axis (β_{ball} and γ_{ball}).

A closer look at Figure 5.5A points out that one of the participants (**s_12**) seems to behave differently from the rest. To illustrate in which sense the behaviour of **s_12** is different from that of the average subject,

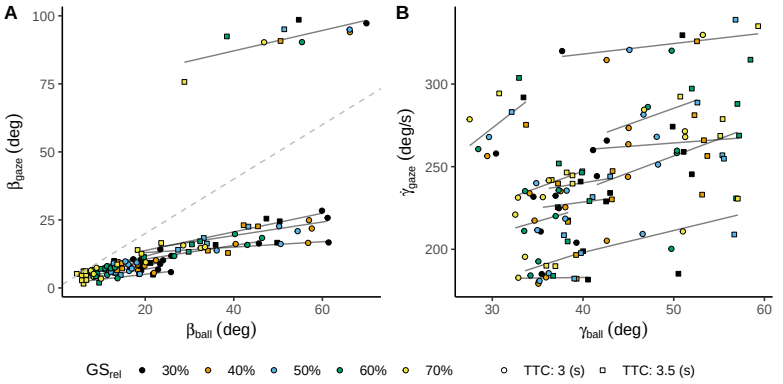

Figure 5.5. A) Gaze horizontal position as a function of ball's position at reappearance. B) Gaze rate of change in the vertical axis as a function of ball's position in the vertical axis. Both panels depict angular measures. Each trajectory (GS_{rel}) is indicated with a different colour whereas shape indicates flight duration (TTC). The grey lines denote the best linear fit per participant. The grey dashed line in Panel A represents the identity line.

we produced Figure 5.6. Figure 5.6 represents how ball's position and gaze direction unfolds across time for the average participant (**s_6**) and **s_12** The average participant (**s_6**) keeps the horizontal gaze direction (β_{gaze}) fixed until the participant decides to look for the ball. From that moment on, the participant tries to keep the ball centred horizontally for the rest of the trial. In contrast, **s_12** rotates horizontally close to the position where the ball will fall at eye-level while the ball is still occluded. Then prior to ball's reappearance **s_12** predicts ball's current position and try to keep the ball centred horizontally during the visibility window.

5.3.2 When do participants prefer to look at the ball?

Do different trajectories influence the moment at which the observers prefer to look for the ball? To analyse if this were the case, we tested whether our participants would change their preferred viewing time (from now on t_{Visible}) across flight duration (TTC) and trajectory (GS_{rel}). The preferred viewing time is defined as the first point in time at which the ball reappears from motion onset ($t = 0$). To do this, we fitted a Linear Mixed Model

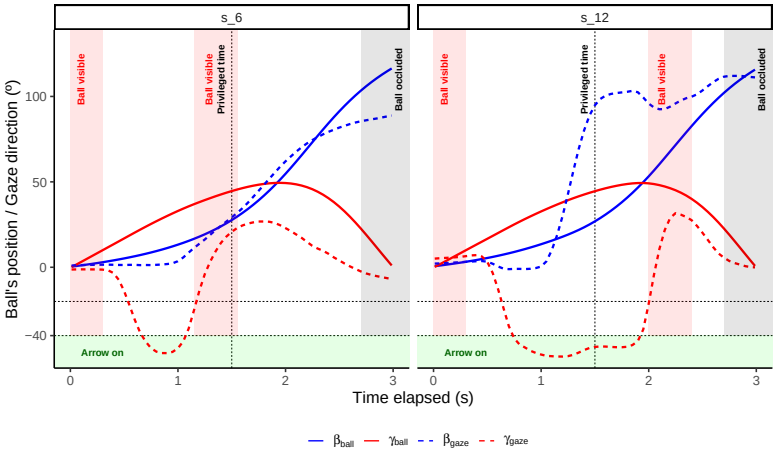

Figure 5.6. Representation of ball's position (solid lines) and gaze direction (dashed lines) in the horizontal (blue) and vertical (red) axis across time for two different participants **s_6** and **s_12**.

predicting the preferred viewing time (t_{Visible}) with trajectory (GS_{rel}) and flight duration (TTC) plus their interaction as predictors (fixed effects). Participant was introduced as random effects.

$$t_{\text{Visible}} \sim TTC \cdot GS_{\text{rel}} + (1|\text{Participant}) \tag{5.6}$$

To test whether our participants exploited different windows depending on the trajectory, we compared the Test model with a Null model only including the random term:

$$t_{\text{Visible}} \sim (1|\text{Participant}) \tag{5.7}$$

A Likelihood Ratio Test proved that the Test model fits significantly better than the Null model ($\chi^2(3) = 65.68$, p <0.001, BIC_{Null}= -2763, BIC_{Test}= -2802). These results indicate that our participants decided when to look for the ball depending on the conditions present at launch (see Figure 5.7A).

Our results point out that the trajectory influenced when the participants preferred to look for the ball (GS_{rel} : Coef. = 0.118, SE = 0.017, $CI_{95\%}$ =

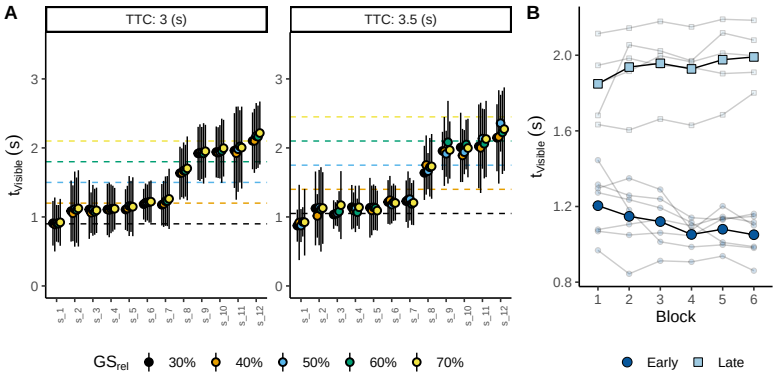

Figure 5.7. A) Mean preferred viewing time (t_{Visible}) per flight duration (TTC in different panels), participant (x axis) and trajectory (GS_{rel} in different colours). The horizontal dashed lines indicate the privileged time points specified by the GS model. The error bars denote $\pm$ a standard deviation. B) Mean preferred viewing time (t_{Visible}) across blocks per participant (small dots) and group of participants (big dots).

0.085 – 0.151, t = 6.872). However, neither flight duration (TTC: Coef. = 0.029, SE = 0.022, $CI_{95\%}$ = -0.014 – 0.072, t = 1.308) nor the interaction term ($TTC \times GS_{\text{rel}}$: Coef. = -0.016, SE = 0.042, $CI_{95\%}$ = -0.098 – 0.066, t = -0.383) reached significance. Hence, our results indicate that our participants adjusted the preferred viewing time to the trajectory but not to the duration of the flight, which is not unexpected since one value of flight duration (3 seconds) was more frequent than the other (3.5 seconds) on a ratio of 5 trials against 1.

As the reader could recall, the GS model allows for accurate estimates of the remaining TTC for the trajectories present in this experiment at different points in time. Specifically, at time points from 30% to 70% (by steps of 10%) of elapsed flight duration. If our participants were exploiting these privileged time points (dashed lines in Figure 5.7A), viewing time should change for both: trajectory (GS_{rel}) and flight duration (TTC). Concretely, viewing time, t_{Visible} would be centred around the corresponding (same colour) dashed lines in Figure 5.7A. However, by simple eye inspection of Figure 5.7A one can notice that the observers use relatively fixed viewing windows and quite independent from those predicted by the GS model. This

was so, despite the fact that their variability could have allowed them to discover, at least, two privileged time points per flight time (In Figure 5.7A the error bars indicate $\pm$ a standard deviation).

It could be argued that our observers did not adjust ball's visibility to different flight durations because there was not enough information to predict TTC in the initial period of visibility. At motion onset the rate of change of the elevation angle ($\dot{\gamma}_{\text{ball}}$) is larger for longer flight durations which translates into higher trajectories. Therefore, at ball's reappearance the rate of change in vertical gaze direction ($\dot{\gamma}_{gaze}$) should be different between flight durations, and this was exactly the case (t (2010.6) = -4.612, $p < .001$). Therefore, our observers predicted differences in ball's vertical position across different flight durations. This result points out that the information needed to estimate different flight durations was available from motion onset.

Further exploratory analysis showed that our participants could be divided into two different groups. Those who looked at the ball before 1.5 seconds since motion onset (*early viewers*; from **s_1** to **s_7**) and those who looked after 1.5 seconds (*late viewers*; from **s_8** to **s_12**). For the sake of simplicity, participants' labels were ordered according to the average time at which the ball reappeared, with **s_1** preferring to see the ball earlier and **s_12** preferring to see the ball later on average. In this regard, a visual comparison across blocks seems to indicate that the participants changed the viewing time throughout the experiment emphasizing the differences between the two groups (see Figure 5.7B). These results are consistent with previous studies in which the observer is free to choose when to gather visual information. Under these conditions, an observer usually explores different viewing windows but ends up exploiting a fixed visibility time window to solve the task (López-Moliner & Brenner, 2016).

5.3.3 Is response time affected by different flight times?

We have just shown that participants can discriminate both flight durations with the initial information at ball's launch. However, our participants used a relatively fixed viewing time to see the ball. In the same line, we should test whether their temporal responses could be explained by using some constant criterion, mapping their actions with a fixed time to solve the task.

This simple strategy, however, would result in the same response time for both flight times. Instead, if they were using some prediction based on the last moments of visibility, we would expect different response times for the two flight durations. To test this hypothesis, we fitted a Linear Mixed Model predicting $\text{Response}_{\text{Time}}$ with flight time (TTC) as predictor (fixed effect) and participant as random effects.

$$\text{Response}_{\text{Time}} \sim TTC + (1|\text{Participant}) \tag{5.8}$$

Then, we compared the previous model with a Null model only including the random term:

$$\text{Response}_{\text{Time}} \sim (1|\text{Participant}) \tag{5.9}$$

The Test Model proved to fit significantly better to our experimental results than the Null model ($\chi^2(1) = 1922.96$, $p < 0.001$, $BIC_{\text{Null}} = 2402$, $BIC_{\text{Test}} = 488$). The intercept for the Test model is 2.988 ($SE = 0.018$, $CI_{95\%} = 2.954 - 3.023$, $t = 162.69$) and the coefficient for TTC, that is, the difference between mean response time for trajectories of 3 and 3.5 seconds is 0.34 ($SE = 0.007$, $CI_{95\%} = 0.326 - 0.354$, $t = 46.52$). Figure 5.8A depicts mean response time per participant and flight time (different shapes). These results show that our participants were able to use different predictive information for the two trajectories. But which sources of visual information were available to perform this prediction?

To answer the previous question, we computed the average difference for each optic cue between both flight durations in the last frame the ball was visible. Then, we obtained the proportion of change per optic cue and inferred whether an observer could use such differences to discriminate both flight durations (TTC). Previous studies have shown that human sensitivity to changes in retinal size (θ), elevation angle (γ) and the rate of change of the elevation angle ($\dot{\gamma}$) is very high, discriminating differences of $\approx$ 5% (Weber fraction) (McKee, 1981; McKee & Welch, 1992; Regan & Kaushal, 1994). Thus, we use this value as a threshold to infer if each visual cue might be readily available to be used for prediction. Our results, as depicted in Figure 5.8B, indicate that all the optic cues considered were discriminable between flight durations. In this sense, since all the variables

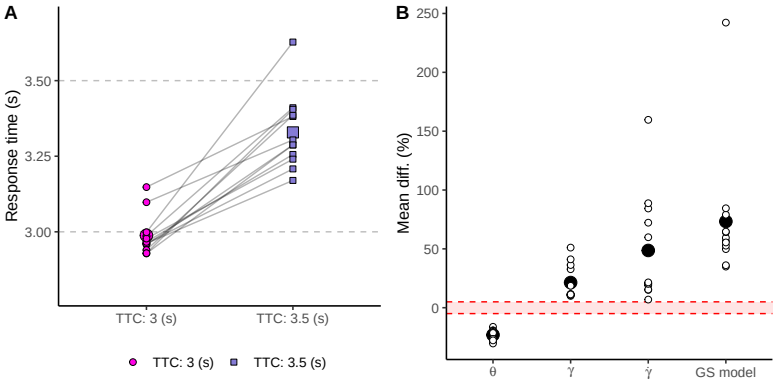

Figure 5.8. A) Response time as a function of flight duration (colour and shape code). Big dots indicate the average across participants. Smaller dots indicate the average per participant. B) Mean difference in percentage between flight durations (TTC) per visual cue and the combination of cues (GS model). The red area denotes differences that cannot be discriminable (lower than $\pm$ 5%). Big black dots represent the average difference across participants. Small dots represent the average difference per participant.

included in the GS model can be discriminated, our results sustain the possibility of combining optic variables in the form of the GS model, which as well, results in a discriminable output for all participants. But how well did early and late viewers estimate the remaining flight time?

For both flight durations, late viewers reflect a better performance in terms of hit rate (Figure 5.9A). However early viewers are more accurate estimating 3 second trajectories (Figure 5.9B). This seems rather contradictory. Thus, what could explain this pattern? Besides the mean error, we also need to take into account the precision in the response (error variability). Figure 5.9C shows hit probability against the standard deviation (SD) of response time. The most precise participants were also the most successful for trajectories of 3 seconds. This can be explained by the decay of prediction on the response variability (de la Malla & López-Moliner, 2015). To confirm this possibility, Figure 5.9D displays the hit probability as a function of the response time since the time of prediction. We defined the time of prediction as the midpoint within the visibility window at which the observer preferred to look at the ball. For 3 second trajectories, those

who relied on longer predictions also present the lower hit probability. Therefore, the prediction decay may explain why late viewers present a higher hit probability.

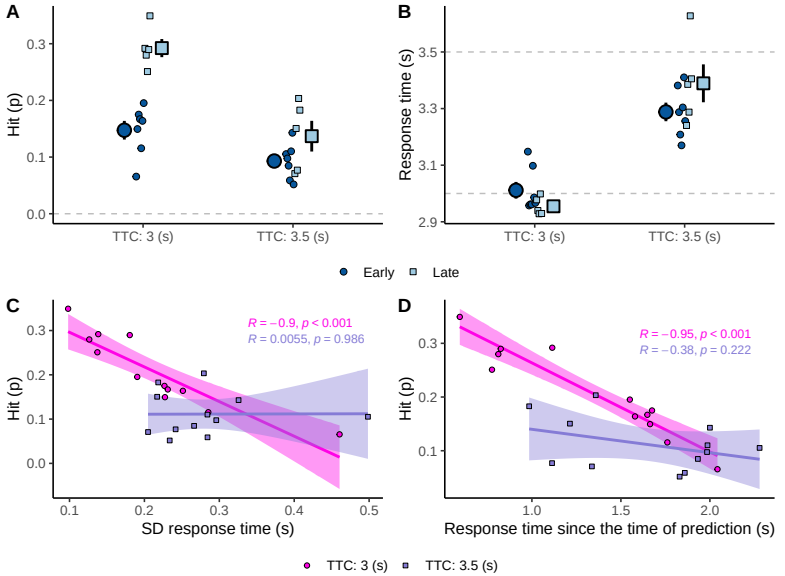

Figure 5.9. A) Hit probability per flight duration (TTC) and group of participants. B) Response time per flight duration (TTC) and group of participants. In panels A and B, the big dots indicate average per group of participants and the smaller dots indicates the average per participant. C) Hit probability as a function of the SD of the response time per participant and flight duration (TTC). D) Hit probability as a function of the response time since the time of prediction per participant and flight duration (TTC).

For 3.5 second trajectories, most participants underestimate the remaining TTC (Figure 5.9B). This underestimation can be explained by the mean response time being pulled down towards the more frequent TTC. In this case, the early viewers seem to underestimate more the remaining TTC than late viewers explaining their performance in terms of hit probability for trajectories of 3.5 seconds (Figure 5.9A).

5.3.4 Are their estimates consistent with the use of information in the GS model?

Our participants seem to be using relatively fixed temporal windows to look for the ball instead of the privileged viewing times specified by the GS model. Besides that, the predictive capacities of the model allow us to study the correspondence with our participants' responses. Figure 5.10A reminds us of the model predictions; the predicted remaining flight time as a function of the time elapsed since motion onset. In the example present in Figure 5.10A, the vertical red line denotes the midpoint of the visibility window (≈200 ms. after ball's reappearance). We computed this moment on a trial-to-trial basis as the time at which the observer gather visual information to draw predictions about the remaining flight time.

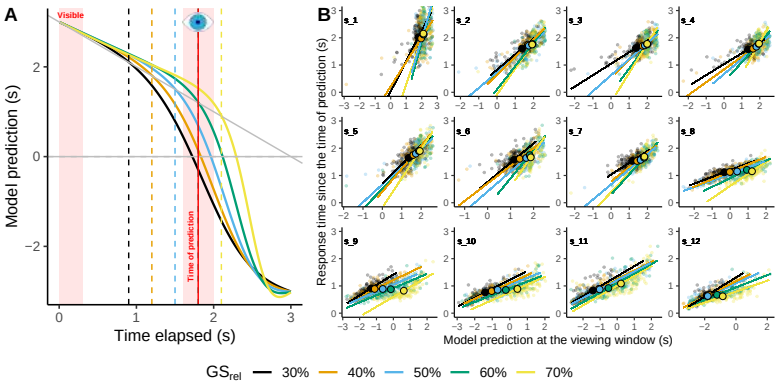

Figure 5.10. A) Predictions of remaining flight time using the GS model as a function of the time elapsed since motion onset and trajectory. The dashed vertical lines indicate the privileged time points specified by the GS model. The vertical red line indicates an example of midpoint visibility time. B) Response time since the time of prediction as a function of the predictions of the model at the viewing window per trajectory (colour code) and participant. The coloured lines indicate single fits per trajectory. Big dots indicate the average per participant and trajectory.

To test if the estimated remaining time at the viewing window correlates with the prediction of the model, we fitted a Deming linear model to the response time since the central moment of the viewing window as a function of the prediction of the model at the same viewing window. We

did so per participant and trajectory (GS_{rel}). Individual fits are represented in Figure 5.10B. The confidence interval for the pooled slopes across participants and trajectories does not contain zero ($CI_{95\%}$: 0.387 - 0.852). This result points out that initially there is a relation between the response time since the viewing window and the predictions of the GS model.

Furthermore, the remaining flight time predicted by the model is different across trajectories: the vertical red line (viewing window) intersects the prediction of the GS model at different predicted values per each trajectory. Concretely, the GS model predicts longer flight durations for those trajectories in which the privileged time takes place later after motion onset (e.g. yellow line in Figure 5.10A). Unlike the previous analysis in which we fitted a linear model within trajectories, we now fitted a linear model across trajectories. We used a linear mixed model with the trajectory (GS_{rel}) as fixed effect (predictor) and participant as random effects. The fitted slope is positive and different from 0 (GS_{rel}: Coef. = 0.322, $SE = 0.025$, $t = 13.111$, $CI_{95\%}$ = 0.273 - 0.371). These results can be inferred in Figure 5.10B, in which the average response time since the viewing window is larger for those trajectories where the privileged time takes place later. However, one participant (**s_9**) does not seem to follow this overall trend (see Figure 5.10B). Because of that, it is important to note here that we cannot rule out the possibility that an observer may be using simpler strategies based on mapping some available optic cues to the temporal response (Zhao & Warren, 2015).

Finally, since the GS model predicts different moments for which the estimation of the remaining flight time is accurate, we decided to check whether when people look at the trajectory closer to those privileged windows leads to some benefit in terms of the temporal error. To do so, we used the median to split the trials per participant and trajectory depending on how close/far they looked at the trajectory with respect to the privileged time specified by the GS model (i.e. 50% closer vs. 50% far). The mean error for the closer trials (-0.031 s.) is smaller than the mean error for the far trials (-0.045 s.)($t(8264.7) = 2.371$, $p = 0.018$). Therefore, these results indicate some benefit for when the observer looks at the trajectory closer to the privileged time points specified by the GS model.

5.4 Discussion

We humans normally track a ball to catch it on the fly. Nevertheless, there are times when this is not possible because we need to divert our gaze somewhere else. By studying the behaviour and performance mimicking these situations, we might gain important insights into how people cope with these scenarios. In this work, we analyse a task in which observers had to judge TTC on the basis of a short periods of visual information. This is a very demanding task in which the participants had to predict ball's position in flight in order to capture the necessary information to estimate the remaining TTC. Our task tries to capture situations in which players control a ball in a parabolic flight while looking elsewhere searching for a teammate.

If possible, an observer will try to reduce the use of an off-line (predictive) strategy by always keeping the ball visible (Brenner & Smeets, 2011; Mazyn et al., 2007). Continuously updating predictions increases reliability (Snowden & Braddick, 1991) and lessens perceptually-driven errors (Belousov et al., 2016; Binaee & Diaz, 2019; de la Malla & López-Moliner, 2015; de la Malla et al., 2018) which explains why our behaviour changes when unexpected perturbations occur (Fink et al., 2009). However, the present task requires a predictive component at all stages.

Congruently with previous literature (Diaz et al., 2013; Jörges & López-Moliner, 2019), our findings show evidence that people can successfully use visual information from a short visibility period at motion onset to predict future balls' positions and likely TTC. After that, restricting mid-flight visibility to a brief window forced our participants to focus on and exploit useful visual information for the estimation of TTC. Previous studies had already shown that it is only necessary to see a short portion of the trajectory to catch a ball (López-Moliner et al., 2010; Whiting & Sharp, 1974). But here we provide further evidence that, there is enough information to estimate TTC despite of the viewing window that is chosen to extract visual information. In this line, the estimations of the remaining flight time are related to the predictions by the GS model and consistently biased across trajectories in the expected direction. However, we cannot rule out the possibility that an observer may be using simpler strategies

based on temporal mappings between different optic cues and the temporal response (Zhao & Warren, 2015).

Gathering visual information in advance can be useful for action planning when a secondary activity has to be performed. However, in our experiment, those participants relying on longer predictions (early viewers) were also more affected by the average flight time and prediction decay. In contrast, those who preferred to see the ball at the very last-minute (late viewers) showed less response variability. This difference seems to be responsible for the better performance of late observers. Nevertheless, despite of the preferred viewing time all our observers preferred to look at the ball in relatively fixed time windows regardless of TTC and the privileged time point specified by the GS model. Why did they decide to look for the ball at those specific moments?

Likely, our participants may not be sensitive to the improvement in hit rate resulting from the use of the visual information within the privileged time points. If the privileged time occurs well before the response, response variability may hinder the detection of an increase in accuracy. Moreover, due to the interleaved presentation of trajectories, our experimental design may have prompted our participants to adopt a unitary strategy to exploit the available visual information rather than taking advantage of the privileged windows specified by the GS model (Amazeen et al., 1999; López-Moliner & Brenner, 2016; López-Moliner et al., 2010). Under this rationale, our observers would be less likely to correct the preferred viewing time. Instead, they would prefer to correct their temporal estimates on a trial-to-trial basis (López-Moliner et al., 2019; Shadmehr et al., 2010).

An important consideration concerning this study is the possible perceptual distortions introduced by immersion in virtual reality (VR) wearing a head mounted display. A number of studies have shown that the perceived space in VR is compressed, presumably due to a conflict between accommodation and disparity signals (Hoffman et al., 2008). This could explain why TTC is consistently underestimated when the estimations are guided by a combination of binocular variables (the rate of change of disparity) and an estimate of distance (Gray & Regan, 1998; Lages, 2006; Rolin et al., 2018). Nevertheless, the geometric layout in our experiment would reduce the impact of spatial compression since distance is uninformative of TTC

in parabolic trajectories with a lateral offset. In line with the GS model, we think that cardinal monocular variables (horizontal and vertical angles: β & γ) are more likely to be used in the estimation of TTC and future positions of the ball. For example, viewing the ball peripherally may bias and reduce the reliability of elevation angle estimates (γ) (Crowell & Banks, 1996) leading to systematic errors in the estimation of the remaining flight time. However, the direction in which visual eccentricity translates into systematic errors might need to be addressed in further studies in which temporal visibility and visual eccentricity can be decoupled.

Tracking a ball visually typically requires the use of both, head and eye movements (Mann et al., 2013). However, in our study, we only have data on head's direction. It is known that eye movements can be quite different from head tracking in interception tasks. Usually, eye movements reflect a predictive component that seems to improve visual pursuit and reflects predictions of a future position of an object at key moments (Diaz et al., 2013; Jörges & López-Moliner, 2019; Mann et al., 2013). An analysis of eye movements could have allowed us to be more precise analysing how the quality of visual tracking improves the estimation of TTC. Furthermore, analysing eye-movements at key moments such as ball's reappearance or the moment the ball falls at eye-level would be interesting to shed light on the predictive component of visual function for the control of interception tasks.

5.5 Conclusion

Our data indicate that the observers are able to use a predictive strategy to estimate both the position of the ball in flight and the remaining flight time using the visual information available during short visibility windows. To estimate the remaining TTC from a mid-flight visibility window, our observers preferred to use fixed temporal windows that might help them to interpret visual information combined with their previous experience.

6 Study 4: Prediction based on a gravity and size prior explains temporal errors and heading locomotion for catching

Catching a flying ball usually requires a locomotion phase towards the interception area and a manual interception phase. An initial prediction of the position and time of interception would allow planning the action overcoming sensorimotor delays and the variability in the visual information related to an observer's movement. In this work, we present a novel formulation in which a spatial-temporal prediction of time-to-contact and interception location is computed combining optic cues and gravitation and ball size knowledge. Previous literature suggested that the interpretation of visual (optical) information would rely on prior knowledge about constants in the world (Earth gravitational acceleration or ball size) to obtain relevant predictions. If such interpretation occurs, consistent deviations are to be expected when the task parameters do not correspond to the ones expected a priori by the observer. In this chapter, we studied if such deviations would be reflected in both phases of ball catching. To do so, we introduced our participants to a virtual environment using a head-mounted display (HTC Vibe @ 90 Hz). A combination of three gravitational accelerations (9.807 m/s^2; $\pm$ 10%), three different ball sizes (0.22 m diameter; $\pm$ 10%) and ten different parabolic trajectories around an observer's initial position were presented. We asked our participants to move as if they were to head the ball in flight. After 90% of the flight time passed, the ball was occluded from view. Then, the observers had to judge the time-to-contact with eye-level using a controller while moving towards the interception location. We show evidence that the temporal errors incurred by our participants correspond to expected from using prior knowledge of gravity and ball size. Furthermore, we found evidence that gravitation had an effect on locomotion congruently with predictions of the interception location.

The present study is being prepared for submission

6.1 Introduction

Catching a ball on the fly is an interception problem that requires two distinct phases. The first is a locomotion phase towards the place where interception is available. The second is a manual interception phase. Even though the first phase has been repeatedly addressed as a case of study known as *the outfielder problem* (Chapman, 1968; Michaels & Oudejans, 1992; Todd, 1981), the influence of the locomotion phase on the visual information sampled to perform the final interceptive action has been largely neglected.

Traditionally, different perspectives would emphasize how an action is controlled based on an online or a model-based control. An online control perspective would envision locomotion and manual interception as independent modules controlled by different sources of sensory information (Fink et al., 2009; McBeath et al., 1995; Peper et al., 1994). In contrast, the model-based perspective would assume that the visual feedback acquired in one phase would serve to refine the action in subsequent phases (Aguado & López-Moliner, 2021a; Kwon et al., 2015) consistently with stochastic control theory (Belousov et al., 2016; de la Malla & López-Moliner, 2015; Todorov, 2004). However, repeated comparisons between both perspectives have shown that neither is superior to the other (Belousov et al., 2016; Höfer et al., 2018). Instead, each one would be able to account for some aspects better.

In line with a model-based perspective, predicting the trajectory followed by the ball (Saxberg, 1987a, 1987b) could allow us to plan our actions. In this respect, there is a debate as to whether reconstructing and predicting a parabolic trajectory is possible or even necessary (Aguado & López-Moliner, 2021a; Fink et al., 2009; Zhao & Warren, 2015). Expert sports players can not describe how visual cues unfold for a parabolic trajectory (Shaffer & McBeath, 2005). Nevertheless, there are sound reasons to assume that such computations may be available for action (Gómez & López-Moliner, 2013; Zago et al., 2009) and, in fact, necessary to avoid sensory-motor delays (Hayhoe et al., 2005; McIntyre et al., 2001; Nijhawan, 1994; Todorov, 2004).

So far, computational models simulated catching based on Cartesian

variables (Belousov et al., 2016; Höfer et al., 2018). However, previously it might be necessary to explore a way to estimate parameters of a task such as the time-to-contact or the interception location based on visual information. Translating retinal information into parameters of the world to act upon is a common problem in vision research (Kersten et al., 2004; Pizlo, 2001). At this respect, the knowledge of different constants in the environment such as such as gravitational acceleration (Jörges & López-Moliner, 2017; McIntyre et al., 2001; Moscatelli et al., 2019), object size (Berkeley, 1709) or even zero-net motion (Stocker & Simoncelli, 2006) have been suggested to calibrate visual information to allow the estimation of spatial and temporal estimates (Aguado & López-Moliner, 2021a; Ittelson, 1951; López-Moliner et al., 2007). However, only the temporal side of the problem has been addressed in the literature (Aguado & López-Moliner, 2021a; Gómez & López-Moliner, 2013). In the following sections, we will describe the necessary computations to estimate the time-to-contact and the interception location in the x-z plane at the eye's height accurately.

6.1.1 Time-to-contact estimation

The GS model (Gómez & López-Moliner, 2013) proposes a solution to estimate the time remaining until the ball returns to eye-level in parabolic trajectories, that is, the time-to-contact. This algorithm relies on a combination of optic variables such as retinal size (θ) and the elevation angle (γ) (see Figure 6.1A) calibrated by internalized *a priori* variables such as Earth's gravity (G) and ball's size (s) in seconds.

$$TTC_{\mathrm{GS}} \approx \frac{2}{G} \cdot \frac{s}{\theta} \cdot \frac{\dot{\gamma}}{cos(\gamma)} \tag{6.1}$$

The correspondence between the a priori assumed variables with the task's parameters assures the accuracy of the predictions. However, if the trajectories' parameters do not correspond to the assumed priors, the resulting calibration would lead to systematic errors. Hence, if either gravitational acceleration or ball size is different from the assumed ones, the predicted time-to-contact would be misestimated. For example, given a constant time-to-contact, the initial vertical speed is larger for greater gravities governing the motion (dark blue parabola in Figure 6.1B). In that

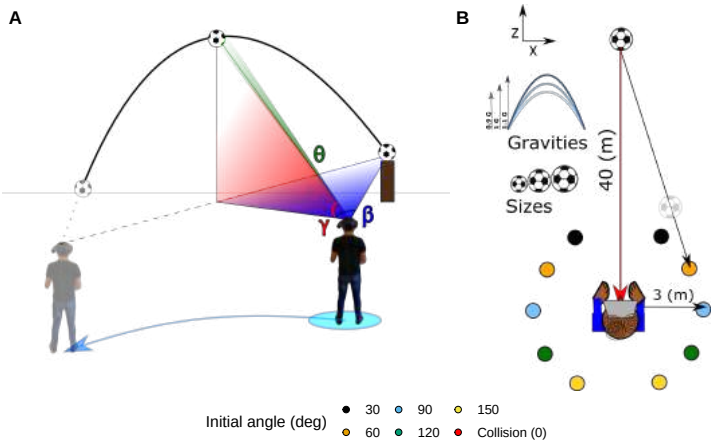

Figure 6.1. A) Representation of the main primitive monocular variables available in a parabolic trajectory: Retinal size (θ), the elevation angle (γ) and bearing angle (β). B) Sketch of the conditions tested in the present work and a collision course trajectory (not tested here). We presented 5 trajectories directed to both sides (dots colour code), 3 gravities and 3 ball sizes.

case, an observer assuming terrestrial gravity would interpret a larger initial vertical speed as if the remaining time-to-contact were greater either. On the other hand, larger ball sizes would have the opposite effect: a general underestimation of the time-to-contact.

6.1.2 Interception location

In this section, we will describe the geometrical specification of the position of a parabolic trajectory in 3D coordinates from a nasal position (between the eyes). Here we will consider x, y and z as the ball's horizontal, vertical and depth position. In Figure 6.2, the reader can find a sketch of lateral view and a bird-eye perspective of a parabolic trajectory with all the necessary angles involved.

Prior to beginning with the spatial formulation, it is necessary to note that we will assume that the observer is looking at the ball at launch time and that the standard ball size and terrestrial gravitation are known to the observer.

With the above considerations, absolute distance would be defined as

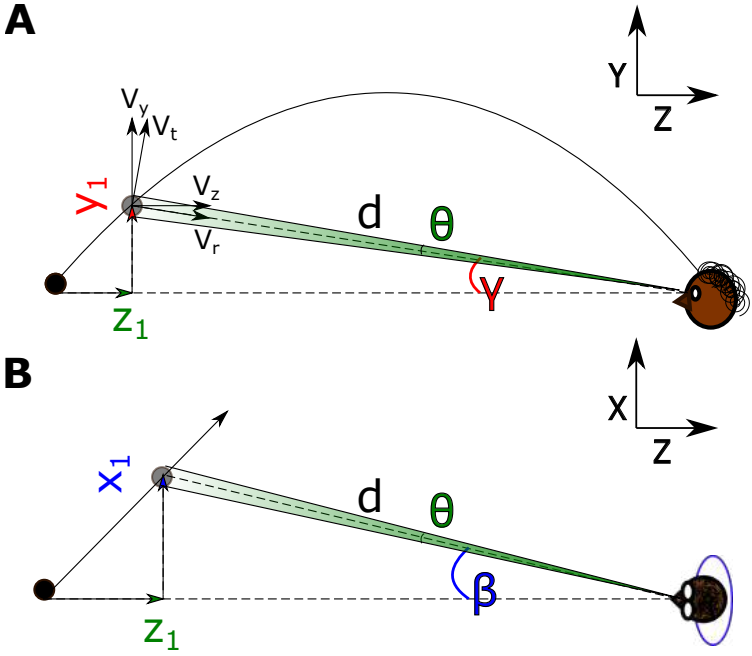

Figure 6.2. A) Lateral representation of a parabolic trajectory. B) Bird-eye view of a parabolic trajectory. Note that the computations in main text assume an estimate of x position at ball's height.

the ratio between the projected retinal size and absolute ball size:

$$d \approx \frac{s}{\theta} \tag{6.2}$$

Following basic trigonometric properties, it is easy to compute the position of the ball in the three-dimensions (see Figure 6.2)[1]:

$$x \approx \frac{s}{\theta} \cdot cos(\gamma) \cdot tan(\beta) = z \cdot tan(\beta) \tag{6.3}$$

$$y \approx \frac{s}{\theta} \cdot sin(\gamma) \tag{6.4}$$

$$z \approx \frac{s}{\theta} \cdot cos(\gamma) \tag{6.5}$$

[1]Note that we estimate lateral ball's position in the plane defined by observer's gaze at ball's height as described by McBeath et al. (2018).

Then, it is easy to see in Figure 6.2 that when t tends to zero, that is, at launch time, the ball's displacement vectors are given by:

$$V_{x0} \approx \frac{\mathrm{s}}{\theta} \cdot \dot{\beta} \tag{6.6}$$

$$V_{y0} \approx \frac{\mathrm{s}}{\theta} \cdot \dot{\gamma} \tag{6.7}$$

$$V_{z0} \approx \frac{\mathrm{s} \cdot \dot{\theta}}{\theta^2} \tag{6.8}$$

However, gravitational acceleration introduces a non-linearity in the projectile's motion, which makes posterior computations more complex. Nevertheless, lateral displacement is still directly measurable if size is known. Its computation is given by:

$$V_x \approx \frac{\mathrm{s}}{\theta} \cdot cos(\gamma) \cdot \dot{\beta} = z \cdot \dot{\beta} \tag{6.9}$$

Estimating both vertical speed (V_y) and depth speed (V_z) is considerably more difficult. To understand why, let us consider depth speed. When an object approaches the observer at a constant speed, depth and radial speed are coupled. Radial speed (see v_t in Figure 6.2) is provided by the first derivative of ball's retinal size:

$$V_r \approx \mathrm{s} \cdot \frac{\dot{\theta}}{\theta^2} \tag{6.10}$$

Thus, depth speed can be estimated from radial speed (V_r). However, in a parabolic flight, the ball does not approach the observer at a constant speed and thus, the lawful relation between depth and radial motion breaks. Radial speed would be still available to the observer. However, like vertical speed, it is no longer constant. Instead, it covariates with a vector tangential to the radial motion, the tangential speed(see V_t in Figure 6.2). Tangential speed envisions the vertical displacement of the ball from the observer's perspective. It can be estimated using the first derivative of the elevation angle with respect to the observer's position:

$$V_t \approx \dot{\gamma} \cdot \frac{\mathrm{s}}{\theta} \tag{6.11}$$

Once radial and tangential speed are recovered, vertical speed (V_y) and depth speed (V_z) can be estimated as follows:

$$V_y \approx V_t \cdot cos(\gamma) - V_r \cdot sin(\gamma) \tag{6.12}$$

$$V_z \approx v_t \cdot sin(\gamma) + v_r \cdot cos(\gamma) \tag{6.13}$$

Hence, it is possible to predict the interception location in the x-z plane:

$$X_{\text{End}} \approx \frac{\text{s}}{\theta} \cdot cos(\gamma) \cdot \dot{\beta} = z \cdot \dot{\beta} \cdot \text{TTC} \tag{6.14}$$

$$Z_{\text{End}} \approx \frac{\text{s}}{\theta} \cdot cos(\gamma) - (v_t \cdot sin(\gamma) + v_r \cdot cos(\gamma)) \cdot \text{TTC} = z - v_z \cdot \text{TTC} \tag{6.15}$$

By predicting the interception location, it would be possible to estimate the angle between the observer and the predicted interception location. Here, we refer to that angle with the Greek letter phi (ϕ).

$$\phi = atan(\frac{X_{\text{End}}}{Z_{\text{End}}}) \tag{6.16}$$

Phi (ϕ) would represent the required angle to reach the interception location. Thus it would correlate strongly with heading angle.

In sum, the above expressions allow us to predict *where* the ball would fall at eye height and the required direction to head towards the interception location using a combination of optic cues with known gravitation and ball size.

Without loss of generalizability, removing ball's size from the above expressions would provide a dimensionless estimate of the interception location. This case will represent a situation where flight duration is known.

$$X_{\text{End}} \propto \frac{cos(\gamma)}{\theta} \cdot \dot{\beta} \cdot \text{TTC} \tag{6.17}$$

$$Z_{\text{End}} \propto \frac{cos(\gamma)}{\theta} - \left(\frac{\dot{\gamma} \cdot sin(\gamma)}{\theta} + \frac{\dot{\theta} \cdot cos(\gamma)}{\theta^2} \right) \cdot \text{TTC} \tag{6.18}$$

Both previous computations would not be useful *per se*. Still, would al-

low to compute the angle towards the interception location if the remaining time-to-contact is known which could be used to guide locomotion.

As a result, an observer might have two different ways of guiding locomotion towards the interception location, either by predicting the interception location or estimating the angle towards the interception location. Assuming that temporal predictions rely on prior knowledge of gravitation and ball size, it is reasonable to assume that both will be essential to predict ball's movement. Hence, if the estimates of the remaining time-to-contact deviate systematically because the assumed priors do not match the parameters governing the trajectory, heading direction would be biased accordingly.

6.1.3 Aim of the study

In this study, we aimed to test whether the predictions of time-to-contact based on the GS model correspond to the estimates provided by our observers under different conditions of gravitational acceleration, ball size and trajectory. Furthermore, the above spatial formulation also provides us with the grounds to compare humans heading towards the interception location with predictions based on optic information and priors. Since the prediction of the interception location relies on an estimation of the remaining time-to-contact it is likely to be misestimated too. Therefore, we predict that the paths travelled across different gravities and ball sizes would reflect those errors. Thus, an observer guiding locomotion would travel different paths towards the interception location between gravitations and ball sizes.

6.2 Methods

6.2.1 Participants

In this experiment, we tested 12 participants (n = 12; 6 self-identified women and 6 self-identified men). One participant had to be discarded due to a particularly noisy eye-tracker's data (filtering procedure removed more than 10% of trials). Participants' age was between 22 and 33 years old with normal or corrected-to-normal vision. All of them were naïve to

experimental goals and volunteered to participate in the experiment. They performed the task wearing a head-mounted display holding a controller with the dominant hand (all were right-handed). The experiment was run by an Intel i7-based PC (Intel, Santa Clara, CA)(i7-9700F). The stimuli were rendered by an NVIDIA GeForce (RTX 2060 SUPER) and sent for display to wireless HTC Vive Pro head-mounted display (HMD) at 90 Hz per eye. Eye movements were recorded using a built-in eye-tracker (Tobii Technology 2011) sampled at 90 Hz.

This work is part of an ongoing research program approved by the local ethics committee of the University of Barcelona in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

6.2.2 Stimuli

We explored ten different trajectories. The interception location was always located 3 meters far from the starting position around the observer describing horizontal angles of 30, 60, 90, 120, 150 degrees to the observer's position either left or right-hand side (see Figure 6.1A).

The ball appeared 40 meters far from the observer's position, vertically aligned to eye level trial-by-trial to account for HMD slippage and posture changes. Flying time was randomly selected from a uniform distribution ranging from 3.15 to 3.85 seconds ($\pm$ 10% of 3.5 seconds). Dynamic effects were neglected.

We used both the standard gravitational acceleration ($9.807 m/s^2$) at sea level and Soccer ball size (0.22 m of diameter) $\pm$ 10% the standard. In total, our experiment consisted of 10 trajectories, three ball sizes and three gravitational accelerations, that is, a total of 90 conditions.

6.2.3 Procedure

Prior to the experimental procedure, the participant and the experimenter tossed back and forth a standard-sized soccer ball to develop familiarity with ball size (diameter 22 cm, Standard size: 5).

Each participant went through a total of 10 blocks of 90 trials each. Each condition was presented once per block. Participants completed 20 training trials before the main experimental procedure to familiarize themselves

with the task and the VR environment. Eye-tracker was calibrated prior to each block and tested using a custom program to ensure that the accuracy of the eye-tracker remained lower than 1.89 degrees of error. The procedure for each trial consisted of the following:

1. The participant was instructed to align both body and gaze towards the ball. Once aligned, a button press using the controller would launch the ball.

2. Once the ball was in the air, the observers had to follow the ball visually while moving towards the interception point.

3. After completing 90% of the flight time, the ball was occluded from vision. The participant was instructed to head towards the position where they would have headed the ball and press a button to estimate time-to-contact. The participants did not receive any feedback on the temporal or spatial response.

The data obtained in this experiment were analysed with R language (R Core Team, 2020) using RStudio (RStudio Team, 2020).

We first ensured that our participants looked at the ball during the whole trajectory as requested. To do so, gaze was categorized as being on the ball if the absolute vertical error to the ball was lower than 6.5 degrees (as done by others Postma et al., 2014) each twentieth of flight time passed. The probability of gaze being on the ball was larger than 90% during the whole time the ball remained visible. The task was self-paced. Each block lasted for about 10-16 minutes.

As a preliminary analysis, we checked that the direction of the ball (left or right) did not affect the perceived time-to-contact. We performed a comparison of mean response time across left-right trajectories. The results indicated that ball direction did not influenced the estimation of flight time ($t(10) = -0.786$, $p = .450$). As a result, we decided to collapse the results for right and left trajectories.

6.3 Results

6.3.1 Temporal estimation

The predictions of time-to-contact based on the GS model revealed the following pattern: an overestimation under gravitations larger than 1G (see Figure 6.3A) and an underestimation for balls bigger than the expected soccer ball (see Figure 6.3B). Furthermore, rearranging the predictions based on the trajectory shows a different pattern of errors depending *mainly* on whether the ball falls behind/ahead an observer's line (see Figure 6.3C) as predicted in previous studies (Aguado & López-Moliner, 2021a, 2021b). See, for example, that the errors are minimal (follow the identity line closely) when the ball falls in line with the observer's perspective (trajectory describing a 90-degree angle with the interception point).

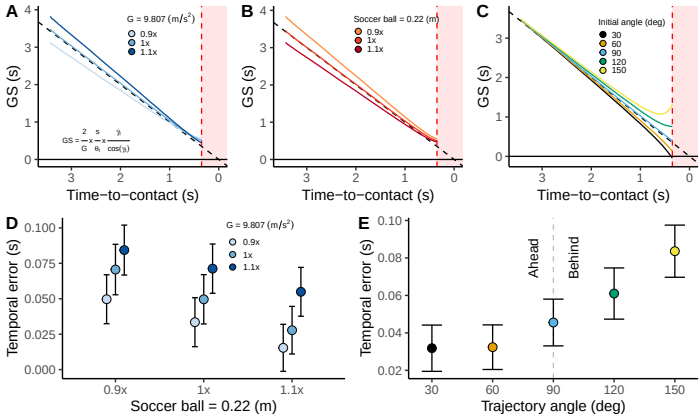

Figure 6.3. Average time-to-contact predicted by the GS model across gravities (A), ball sizes (B) and trajectories (C). The red area denotes the occluded portion of the trajectory. D) Temporal errors per ball size (x-axis) and gravitational acceleration (colour code). E) Temporal errors per trajectory. The error bars indicate a 95% confidence interval in panels D and E.

In line with the previous predictions, our observers overestimated the time-to-contact for trajectories governed by larger gravitational accelerations (see Figure 6.3D; $F(2, 9536) = 13.686$, $p < .001$) while undershooting the time-to-contact the larger the ball size presented (see Figure 6.3D; $F(2, 9536) = 12.269$, $p < .001$). Both trends align with the initial predictions

drawn from using priors to estimate the remaining time-to-contact. Furthermore, when the balls fell behind our participants' line, they tended to overestimate the remaining time-to-contact and vice-versa (see Figure 6.3E; $F(4, 9536) = 11.924$, $p < .001$).

A closer look at the predictions drawn by the GS model across gravities during the trial indicates a pattern of temporal errors inverse to that of the initial predictions (predictions intersect ≈ 0.5 seconds prior to time-to-contact in Figure 6.3A). The observer's locomotion would explain the differences in the optic flow. At the end of the trajectory, the distance between the observer and the ball is relatively short. Therefore, small movements would affect considerably the optic flow rendering optic information unreliable. Thus, it stands to reason to avoid weighting heavily latter information to perform a prediction. Instead, updating the predictions in the light of our prior estimates during the course of the trial would lead to an overall better performance.

Sources of information available

So far, we assumed that our participants solved the temporal task using standard priors of gravity and ball size. However, they could extract the actual parameters from the information contained in the optic flow during the trial. For example, recall that stereoscopic images were provided, and thus, absolute ball size could be recovered (Regan & Beverley, 1979). Another strategy would entail establishing temporal mappings with optic cues or correlates of retinal expansion such as Tau (Lee, 1976). To test this possibility, we studied the correlation between different sources of information and the mean time-to-response across the conditions present in the experiment (trajectory, gravity, and ball size) at different proportions of flight time passed, concretely each twentieth of flight time passed (see Figure 6.4). We defined the time-to-response as the time elapsed from occlusion to the observer's response.

Concretely, we analysed retinal size (θ), Tau (τ), the elevation angle (γ), the rate of change of the elevation angle ($\dot{\gamma}$), the output of the GS model using fixed priors corresponding to the standard values of gravity and size ($GS_{\text{Fixed priors}}$) and the output of the GS model using correct values of G and size as priors ($GS_{\text{Correct priors}}$). Note that the latter variable envisions an

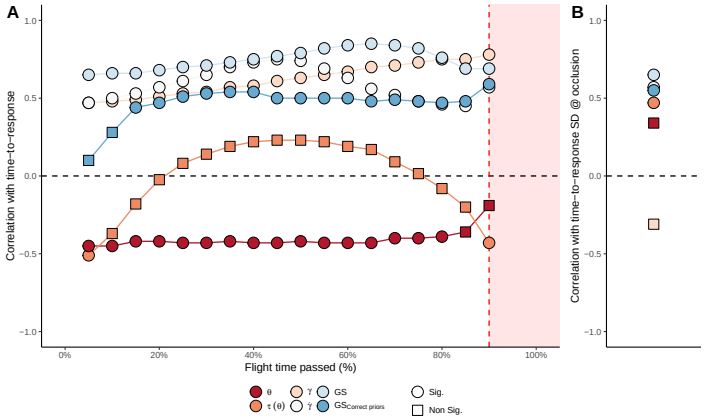

Figure 6.4. Correlation between different variables and the time-to-response, that is, the time elapsed from occlusion to the observer's response. The red area denotes the occluded portion of the trajectory. To account for multiple comparisons (multiple variables), we used an alpha of $0.05/6 = 0.008\overline{3}$.

observer who can accurately extract gravitation and absolute size during the trial.

All variables but retinal size (θ) and Tau (τ) present meaningful and statistically significant correlations with time-to-response at occlusion. Our analysis highlights the average superiority of the GS model based on fixed priors (GS) compared to other sources of information. The output of the GS model using fixed priors reflects the largest correlation with the time-to-response of them all. However, its correlation at occlusion is smaller than that based on a temporal mapping with the elevation angle (γ).

Moreover, we analysed the correlation between the SD of different sources of information and the SD of the response time (see right margin in Figure 6.4). This analysis reflects how likely is the variability in the temporal response given the variability contained in the each piece of sensory information analysed. Retinal size θ, the rate of change in the elevation angle ($\dot{\gamma}$) and the both predictions based the GS model present statistically significant and positive correlations with time-to-response standard deviation. However, despite the strongest correlation with mean response time, the variability of the elevation angle (γ) at occlusion shows no relationship with response variability, which casts doubts on its direct use.

6.3.2 Locomotion phase

In order to analyse the trajectories travelled by our observers we first had to normalize the flight time as the proportion of flight time passed. Then, we split each trial into a hundred steps corresponding where the 100th step would correspond to the time at which the ball returns to eye-level. Then, we summarized the position of the observer in the x-z across step, gravitation, size, angle and mean time-to-contact (3 bins).

To analyse the trajectories across different values of gravitation and size we fitted a smooth local regression predicting the position of the observer in depth (z) from the horizontal position of the observer (x) including the effects of gravity, size and time-to-contact (each independently) and the interaction with the different parabolic trajectories. Then, we applied an ANOVA to analyse individual factor effects explaining variability. We will not report main effects for the different trajectories because the participants had no trouble navigating to the interception point. Thus, trajectory main effects explain a significant portion of variability.

The first thing that stands out when we look at Figure 6.5A is that our participants travelled a different path towards the interception point depending on gravitation ($F(2, 13206.48) = 25.513$, $p < .001$). With a gravity larger than the terrestrial, the paths shift greater away from ball's position in all trajectories. The interaction term did not reached significance ($F(8, 13206.48) = .488$, $p = .866$). This pattern is consistent with the observer predicting that the ball will fly further due to an overestimation of the time-to-contact. Clear examples of this pattern are those trajectories ending behind the initial position of the observer Figure 6.5A. On the contrary, we did not find relevant differences between the paths followed by our participants across ball sizes ($F(2, 13206.48) = 2.143$, $p = .117$; see Figure 6.5B) nor its interaction term with the trajectory ($F(8, 13206.48) = 1.027$, $p = .413$).

In principle, if solely the kinematics of the ball guide locomotion, we might expect similar paths resulting from different gravities and flight times, for which vertical motion correlates. Grouping the trials by bins of mean flight time, we did not found a main effect of mean time-to-contact ($F(2, 13206.47) = 1.784$, $p = .168$). Instead, we found an effect of the interaction between the trajectory and mean time-to-contact ($F(8, 13206.48)$

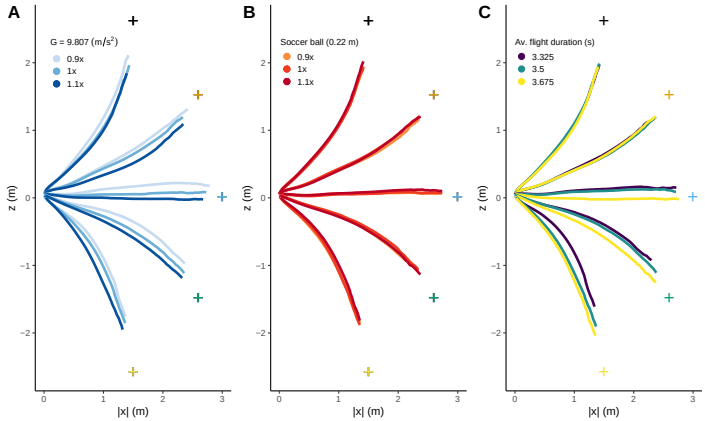

Figure 6.5. Average path followed by the observers across gravitational accelerations (A), ball sizes (B) and mean flight duration (C) for different trajectories. The observer moves from the origin. The ball launches laterally and vertically aligned with the observer at a distance of 40 meters far [$x_0 = 0$,$z_0 = 40$].

= 15.025, $p < .001$) showing that time-to-contact only had an effect for some trajectories. Concretely, between those trajectories where the ball travels to the side or behind the plane of the observer (Figure 6.5C). This dissociation between the kinematics of the ball and the path travelled by our observers supports the idea that prior knowledge of gravity might play a role interpreting visual information to navigate towards the interception location.

Given the above results across flight durations, we decided to analyse movement speed (heading speed) between the different mean flight durations. We did so to find out if there were pre-planned differences in locomotion. To analyse when those differences are relevant, we divided each trajectory into tenths of flight time passed. Then, we applied independent ANOVAs predicting heading speed as a function of average flight duration each tenth of flight time. In Figure 6.6B the reader can see that there are no differences in movement speed during the first half of the trial. Nevertheless, movement speed differed across flight durations at 70% ($F(2, 30) = 6.621$, $p = .004$), 80% ($F(2, 30) = 5.923$, $p = .007$) and 90% ($F(2, 30) = 3.380$, $p = 0.047$) of flight time passed. These results suggest that if locomotion is pre-programmed, the observer would be using an approximate estimation

of time-to-contact based on prior experience from motion onset. Then, as more information is integrated, time-to-contact predictions are refined, allowing the observer to modulate locomotion speed as a function of the predicted remaining flight time. In support of this hypothesis, recall that in the previous section we found that around 65% of flight time elapsed, the correlation between the predictions of the GS model and response time reaches a peak. Combined, both results suggest that temporal predictive information is available.

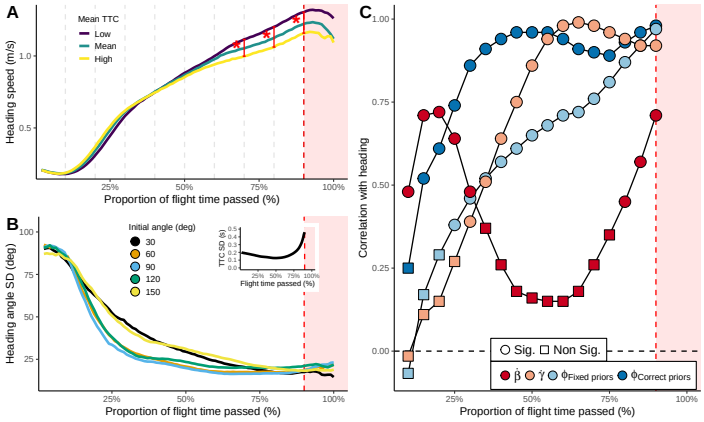

Figure 6.6. A) Average heading speed versus proportion of flight time passed across three groups of flight durations (colour code). B) Average SD of heading versus proportion of flight time passed across trajectories (colour code). The inset represents the average SD of the temporal predictions based on the GS model. C) Correlation between different variables and time-to-response versus proportion of flight time passed. In all panels, the red area denotes the occluded portion of the trajectory. To account for multiple comparisons, we used an alpha of $0.05/4 = 0.0125$.

To further analyse if heading is congruent with refining predictions, we computed the SD of the heading angle, that is, the direction in which the observer moves. The SD of heading would indicate how much the observer relies on online information. Figure 6.6B shows that, at motion onset, heading variability is highest. After around ≈ 0.385 seconds it decreases steeply. Then, for those trajectories falling closer to the observer's plane (yellow and black lines), heading variability decreases slower. But, how different pieces of information contribute to control heading over time?

To answer this question, we performed an analysis correlating variables associated to the continuous control of heading with the average heading angle across trajectories, gravitations and ball sizes (Fink et al., 2009; McBeath et al., 1995; McLeod et al., 2006). Thus, we correlated heading direction with the horizontal and vertical displacement of the ball on the retina ($\dot{\beta}$ and $\dot{\gamma}$ respectively). Furthermore, we included two estimates based on our spatial formulation predicting the necessary heading direction to reach the interception location. Recall that we defined this variable above with the Greek letter phi (ϕ) (see Equation 6.16). The first estimate included only visual cues and a prediction of the remaining time-to-contact using fixed priors of gravitation and ball size (from now on $\phi_{Fixed\ Priors}$). The second used priors of gravitation and ball size adapted trial-to-trial (from now on $\phi_{Correct\ Priors}$). To account for sensorimotor delays accommodating visual information to heading angle control, we performed a correlation with the above computed variables delayed ≈ 0.315 s., that is, a 10% of the average portion of the trajectory visible (Belousov et al., 2016).

At first glance, we can see in Figure 6.6C that horizontal and vertical displacement correlate with the heading angle at different portions of flight time. This is an expected result, as the typical motion pattern reflects a more pronounced lateral component in the displacement at onset and then a depth component more reliant on the rate of change of the elevation angle. Estimates based in $\phi_{Fixed\ Priors}$, present an increased correlation as the time-to-contact approaches, increasing sharply from 70% of flight time passed. Moreover, the estimates based on the correct priors ($\phi_{Correct\ Priors}$) correlate strongly with the direction of heading over the entire trajectory. This result was expected since our observers are, in fact, heading towards the interception location and this algorithm signals accurately the heading direction required.

Putting together the results in Figure 6.6, it is interesting to note that the larger the correlation between heading and the predicted heading angle with fixed priors, the larger the difference in heading speed. Also, note that at that moment the correlation between the predicted time-to-contact using the GS model and the time-to-response reached a peak in Figure 6.4. Taken together, these results may show evidences of an initial use of a rough mean estimate of the remaining time-to-contact at the beginning of the trial to

guide heading. Then, the observers would adapt locomotion continuously as new evidence is integrated.

6.4 Discussion

6.4.1 Temporal estimation

Our results show that the temporal errors incurred by our participants are coherent with the use of gravitational acceleration and ball size priors, this is shown in Figure 6.3. Specifically, we found that the predictive information in the GS model using internalized priors explains the temporal estimates provided by our observers better than the prospective temporal information from Tau, the use of temporal mappings with optic cues (Aguado & López-Moliner, 2021b; Zhao & Warren, 2015) or computations based on the GS model with correct priors adapted trial-by-trial.

First, Tau (τ) shows a negative correlation with the estimated time-to-contact (see Figure 6.4). Second, the use of temporal mappings based on the rate of change of the elevation angle ($\dot{\gamma}$) would be available to solve the present task. However, we consider that its application would not be consistent with the variability present in real life. Instead, it would require contexts where the visual conditions are relatively constant, just like experimental designs. Some evidence against the use of $\dot{\gamma}$ may come from previous studies where time-to-contact could be estimated even when the observers only viewed the first half of the trajectory (Aguado & López-Moliner, 2021b; Amazeen et al., 1999; López-Moliner & Brenner, 2016). Third, we showed that the predictions made by our participants conformed to our initial predictions based on fixed priors. In fact, our analysis shows that using fixed priors fits better overall than using priors adapted trial-by-trial.

Note, however, that the instantaneous temporal errors predicted by the GS model at occlusion did not correspond to the errors incurred by our participants. Such pattern led us to question why this would be the case. We reasoned that using the information contained in the GS model would benefit from using fixed temporal windows in which prediction is most reliable, weighting more mid-flight visual information than information at

the end of the trajectory.

6.4.2 Locomotion phase

We found that observers modulate their trajectory towards the interception point depending on the gravitational acceleration governing the ball's kinematics (see Figure 6.5). Large gravitational accelerations prompt higher vertical speeds given a constant time-to-contact. Thus, unexpected large gravities could have led our observers to predict greater distances travelled by the ball, which, in turn, may have biased our participant's locomotion. From an online control strategy (McBeath et al., 1995; Oudejans et al., 1997; Zhao & Warren, 2015) one could argue that those differences can be explained by the differences in the kinematics of the trajectory. Nevertheless, we found differences in the paths travelled across different gravitational accelerations not present between different flight durations. This dissociation casts doubts on the use of correlates of optic cues signalling the vertical displacement of the ball. Instead, it would be more congruent with an explanation based on using a prior of gravitational acceleration to interpret sensory information.

In contrast, we found no an effect of ball size for locomotion. Likely ball's size could have been derived from binocular information (Berkeley, 1709; Regan & Beverley, 1979), realizing that the initial distance was constant (Hecht et al., 1996; McConnell et al., 1998; J. Watson et al., 1992) or a combination of both. Extracting absolute ball size would remove the effect of assuming a single ball size during locomotion when visual information is available. At this respect, previous studies suggest that known ball size might not be considered when visual information is available (Armbrüster et al., 2008; Hu & Goodale, 2000) or when other pictorial cues are missing (Rushton & Duke, 2009). Our spatial formulation would only rely on known ball size to predict the remaining time-to-contact. If time-to-contact is already known, visual information would suffice to guide locomotion towards the interception location successfully.

At this respect, it is interesting to note the temporal coupling between the following: (1) the maximum correlation of the predictions of the GS model with the estimates of time-to-contact (see Figure 6.4) (2) the change in speed across flight durations (see Figure 6.6A) and (3) correlation of

the heading angle with the spatial predictions using fixed priors (see Figure 6.6C). The fact that the three are coupled seems in agreement with an explanation based on computations of the interception location based on known time-to-contact. That prediction would signal *about* the right heading direction from motion onset. In this sense, first the observer might prefer to accommodate lateral displacements explaining the initial decrease in heading variability (see Figure 6.6B). Then, as further temporal and spatial predictions refine, prediction might take over the control of heading.

In line with a predictive mechanism, we identified different reductions on heading variability across the trajectories tested. For example, the more aligned the trajectory with the observer's sagittal plane, the lower the decrease of heading variability. These results show that an observer's movement in depth is generally more variable than its lateral counterpart, which indicates that both might rely on different sources of visual information. Our spatial formulation indicates that there may be two main sources of error in calculating the interception point in depth: estimates of tangential velocity, derived from the vertical displacement of the projection of the ball on the retina, and estimates of radial velocity, derived from estimates of the visual expansion of the object on the retina. Tangential speed is the same across trajectories at motion onset. Therefore, it is not a likely candidate to be the source of heading variability. On the other hand, radial speed is different across the trajectories tested at motion onset. Radial speed is based *mainly* on retinal expansion, which would be hard to discriminate at large distances (Gómez & López-Moliner, 2013; J. Harris & Watamaniuk, 1995; Regan & Beverley, 1978). And thus, might explain heading variability decreases for trajectories close to the observer's plane.

In sum, the above results highlight the open-loop nature of locomotion and the possibility of modelling catching behaviour (Belousov et al., 2016; Höfer et al., 2018). Specifically, our spatial formulation based on optic cues provides the grounds to compare human catching accuracy and precision with predictions based on tools such as Kalman filters (Kalman, 1960) that integrate predictions with psychophysically plausible precision levels across time.

At this point, we would like to mention some of the limiting factors in interpreting the results of this study. First and foremost is the use of a

head-mounted display. Within an HMD, some visual cues such as disparity or accommodation are often in dispute, which makes generalizing our data to real-life problematic (Bingham et al., 2001). In order to reduce the effect of this mismatch, further studies would benefit from letting the participants calibrate the visual space interacting with the environment prior to the experimental sessions (Hornsey & Hibbard, 2021; Kelly et al., 2013).

In this study we did not render dynamic effects such as air resistance. Air resistance would have affected the ball's kinematics differently depending on ball's size (Aguado & López-Moliner, 2021a; Brancazio, 1985). Thus, if considered, an observer's prediction of the interception location would have been different across ball sizes. Still, we found no differences across the paths travelled. This result suggests that an observer's computations would not include either air resistance or even known ball size at all to guide reaching (Peper et al., 1994). Still, the relevance of prior knowledge of ball size remains to be explored in further studies.

In the present experiment the participants were asked to estimate the time at which the ball would fall at eye level. Although they reported doing so, the results in Figure 6.3 show a general overestimation of the time-to-contact. In our view, this mismatch could be explained by the fact that the temporal response was provided by using a hand-held controller, which may have biased the position at which they were estimating the time-to-contact.

6.4.3 Conclusion

The present study shows that both phases in the interception of a ball in parabolic flight are likely to be affected by prior knowledge about a flying ball and the variability of visual information. Our results show differences in the estimation of the remaining time-to-contact of a flying object consistently with using both gravity and size priors. Furthermore, we found that our observers modify their path towards the interception location based on known gravitation and the estimated time-to-contact. Concretely, we found that the initial locomotion is more consistent with using a rough estimate of time-to-contact based on prior knowledge of flight time. Then, as time goes by, the predictions of remaining time-to-contact become more relevant to guide locomotion congruently with a continuously updated

prediction.

7 General Discussion

At the beginning of this dissertation, we introduced the inverse problem of Perception and outlined different ways of addressing it. In this work, we took the stance of a Bayesian perspective based on knowledge about statistical regularities of the world. Those regularities would serve as constraints helping an observer infer features of the world correctly or draw predictions. However, an overgeneralisation of those statistical regularities in the inferential process would lead to systematic errors.

This dissertation uses those systematic errors to compare the predictions made by different models based on a priori knowledge with human behaviour. In Study 1, we analysed if the assumption of a stationary world could explain systematic errors estimating the direction of an object moving in depth. Then, in Study 2, 3 and 4, we used a model that relies on a priori knowledge of terrestrial gravitation and ball's size to draw predictions of the remaining time-to-contact. Then, we compared our predictions with our participants' to investigate their correspondence.

7.1 Is motion perception guided by prior expectations?

7.1.1 Motion estimation

Under the perspective taken in this thesis, our prior knowledge of the world modulates our expectations about the dynamics of the objects around us. Concretely, specific contexts activate radically different expectations about the objects around us. For example, the expectation we have about the movement of a car when crossing the road is not the same as the

expectation of a free-falling ball. While the car will decelerate, the ball will accelerate. Previous literature showed that simple associative mechanisms might suffice to learn new sensorimotor mappings (Bedford, 1989) even when the discriminants of the new mapping are not directly related to the task (van Dam & Ernst, 2015). Thus, it might be possible to learn that cars decelerate when a pedestrian approaches and that a ball will accelerate when viewed in the air.

In Study 1, in turn, the expectation that the objects around us remain still is key. The studies of Thompson (1982) and Weiss et al. (2002) highlighted a general tendency to underestimate the local motion of a stimulus with low contrast. They assumed that the lower the reliability of visual evidence, the larger the attraction towards the prior in the posterior estimate. The mechanism underlying this bias was then proposed by Welchman et al. (2008) to explain why judging the direction of objects moving in 3D is usually biased. In theory, estimates of motion in depth are less reliable than estimates of lateral motion. Thus, motion-in-depth estimates would be more attracted to a slow-motion prior than lateral motion within a Bayesian model. This differential attraction towards the prior would result in a bias in the perceived direction of movement. Indeed, in Study 1, we showed that our participants overestimated those trajectories displayed at higher speeds congruently with motion estimates following Weber's law.

So far, Bayesian models and, concretely, the slow-motion prior had been criticised for being loosely specified (Bowers & Davis, 2012). Nevertheless, our model was constrained by just one free parameter, the variability of the slow-motion prior. This finding is a hallmark because our model provided a coherent explanation of how directional judgements change at different speeds, and thus, it is likely generalisable to any movement speed. Other criticisms against the slow-motion prior point out that the perceived speed of a low-contrast stimulus can also be overestimated (Bowers & Davis, 2012; Rahnev & Denison, 2018; Thompson et al., 2006). In this regard, Moscatelli et al. (2019) shows an example where the speed of a stimulus moving downwards is underestimated at low contrast and overestimated at high contrast. They argue that the overestimation would be congruent with an expectation of gravitational acceleration. Motion overestimation would only apply to high contrast stimuli similar to the objects we interact with

daily. This example would illustrate how different contexts would prompt radically divergent expectations. Thus, understanding the inferred causes of the observed movement would allow us to understand an observer's interpretation, not only quantitatively but also qualitatively.

7.1.2 Motion extrapolation

When our task is to intercept an object or acknowledge when it will arrive at a pre-specified location, extrapolating the stimulus's movement is vital. Motion extrapolation is crucial because the sensory system has to cope with systematic delays when processing sensory input and therefore cannot make use of instantaneous information (Nijhawan, 1994). For example, predictive mechanisms based on motion extrapolation would be needed to explain the "flash-lag" effect, a visual illusion in which a moving stimulus is perceived as ahead of a stimulus that flashes unexpectedly in the same position (Khoei et al., 2017).

However, for a 3D parabolic trajectory, extrapolating motion is not sufficient to accurately predict when a ball will fall at eye level. Take, for example, our "early viewers" in Study 3. Those participants only viewed portions of the trajectories where the elevation angle increased in combination with an expansion rate irreconcilable with their temporal estimates. Under those conditions, they would have no way of extrapolating 3D motion to estimate a time-to-contact. Nonetheless, our participants predicted the remaining time-to-contact with relative accuracy. Another available strategy that would override the need to extrapolate motion could be establishing temporal mappings with different optic cues. In Study 3, we could not discard their use. However, in Study 4, we were able to discard the use of temporal mappings with variables such as retinal size, the elevation angle or Tau (Lee et al., 1983; Zhao & Warren, 2015). The only common available optic cue to establish a temporal mapping would be the rate of change of the elevation angle. However, even in that case, an expectation gravitational acceleration would be necessary to estimate the remaining time-to-contact accurately (Brenner et al., 2014; Brenner & Smeets, 2018).

As discussed in Study 2, we are rather bad at discriminating arbitrary accelerations (Gottsdanker et al., 1961b; Werkhoven et al., 1992). However,

some studies showed that given enough time, we might be able to partially attune motion extrapolation to arbitrary accelerations (Bennett & Benguigui, 2013; Bennett et al., 2007). According to Jörges and López-Moliner (2017), the assumption of a terrestrial gravity prior would facilitate the estimation and prediction of motion for gravitationally accelerated objects. Indeed, when faced with a task such as catching or tracking a ball on the fly, we seem exceptionally attuned to trajectories under terrestrial gravitational acceleration. Jörges and López-Moliner (2019) showed that observers track more accurately trajectories in the frontotemporal plane at 1g than equivalent trajectories at -1g. As previously shown by others (Postma et al., 2014) in Study 4, we checked that our participants had no problems tracking 3D parabolic trajectories. Furthermore, in Study 3, we found that our observers modulated the speed at which they looked up, trying to find the ball depending on the initial vertical motion. These results support the assumption that observers can use vertical visual speed and use it to predict the position of the ball even if they cannot deliberately describe its visual pattern (Reed et al., 2010; Shaffer & McBeath, 2005).

At this point, we would like to draw the reader's attention to the spatial formulation developed in Study 4 for the outfielder problem. Our algorithm describes how to predict the interception point or the necessary heading angle towards the interception location based on a combination of optic cues and prior knowledge in a similar way as the GS model (Gómez & López-Moliner, 2013). To our knowledge, this is the first time that an algorithm to predict the interception location in Cartesian units has been put forward for parabolic trajectories using optic cues. In this respect, Study 4 provides some insights into how optical information is processed to extrapolate motion-in-depth. We found that our observers travelled different trajectories depending on the gravitational acceleration present. Remember that gravitation only appears in the estimation of the remaining time-to-contact in our formulation. Therefore, the differences in the paths travelled suggest a computation of motion-in-depth and an extrapolation in time that incorporates a gravity prior. On the other hand, we did not find differences between the trajectories followed by our observers when exposed to different ball sizes. Those results may reflect the challenges faced by the system trying to incorporate pictorial cues to calibrate optic

information (Hornsey & Hibbard, 2021; Rushton & Duke, 2009). If computations based on known ball's size are not available mid-flight, relying on prior knowledge of time-to-contact, as suggested in Study 4, would be helpful to guide action until visual information increases reliability.

7.2 Is time-to-contact estimation guided by previous expectations?

In principle, to estimate the time-to-contact with a predetermined location, it would be necessary to calculate both object's position and speed. As we have seen above, estimates of an object's motion can be affected by different a priori variables, either the slow-motion prior or the expectation that an object will move in an accelerated way due to gravitational acceleration.

Concerning the gravitational acceleration, a terrestrial gravitation expectation has been suggested for accurate timed-actions in free-fall (Lacquaniti & Maioli, 1989) or parabolic trajectories in the fronto-parallel plane (Jörges & López-Moliner, 2019, 2020a) or a head-on approach (de la Malla & López-Moliner, 2015).In the last decades, candidate regions for the processing and localisation of a gravitational prior have been described in the literature (Indovina, 2005; Miller et al., 2008). On its part, knowledge of ball size has been suggested to be critical in different time-dependent tasks such as judging the time-to-contact (DeLucia, 2005), establishing a starting time for the initialisation of an interceptive action (López-Moliner et al., 2007) or even judging the distance at which an object will pass by us (Battaglia et al., 2005). Furthermore, previous studies have shown that time-to-contact estimates can be modulated by other factors such as the variability in the size of the balls presented (López-Moliner & Keil, 2012) or even perceived threat (Vagnoni et al., 2015).

In Study 2, we put forward a biologically plausible neural mechanism that incorporates knowledge of the ball's size and terrestrial gravitation. In our simulations, we show how known gravitation and ball size reduce the ambiguity in the interpretation of visual information. The result is an algorithm that allows predicting the remaining time-to-contact accurately. In Studies 3 and 4, we found pieces of evidence of computation of the time-to-contact in the direction predicted by the GS model across different

gravitational accelerations and ball sizes. After discarding alternative explanations based on temporal mappings, the alternative that gathered the most convincing evidence was a computation in the form of the GS model. Thus, relying on priors of terrestrial gravity and standard ball size even over a variant where the correct gravity and ball size were known to the observer. In both studies, we go a step further testing the prediction of time-to-contact for 3D trajectories in a non-collision course with the observer. Testing different trajectories allowed us to compare the errors committed by our participants with the systematic errors predicted by the GS model when the ball is not in a collision course with the observer. Our results matched the GS model predictions: overestimating the remaining time-to-contact for trajectories falling behind the observer's position and vice-versa.

It is essential to mention here that gravitation or ball size are not the only internalised pieces of information available for an observer to estimate the remaining time-to-contact. For example, in Study 3, our observers preferred to look at the ball during a constant time window. This strategy may be beneficial because it reduces the variability of the visual information extracted from the optic flow to estimate the time-to-contact. Thus, it allows for correcting future estimates based on previous errors. In the same study, we presented a reduced number of catch trials with a flight time slightly longer than the standard (500 ms). In those trials, time-to-contact was systematically underestimated. Nevertheless, their errors across trajectories were consistent with the predictions from the GS model. These results suggest that our participants incorporated information from previous trials to weight optic information.

7.3 Have our goals been achieved?

The main objective of this thesis was to study how priors affect the processing of visual information to provide a solution to the inverse problem. During the introductory chapter, we framed the scope of the thesis focusing on how motion and time-to-contact are estimated in combination with internalised prior knowledge.

Our **first objective** was to analyse whether direction estimation derives

from motion estimates based on a combination of visual and prior information. Study 1 showed that a Bayesian model based on motion estimation with a preference for slow speeds sufficed to characterise 3D direction judgements. This study showed that direction judgements are biased as a function of motion speed combined with a single prior towards low speeds.

The scope of the **second objective** was more general: to analyse the existing evidence of using priors in the form of knowledge about gravitational acceleration and object size to estimate the time-to-contact. Study 2 reviewed the literature supporting the relevance of gravity and ball size priors. We showed that both might be important in coordinating depth judgements, providing velocity constancy and guiding sensorimotor tasks (e.g. catching or tracking). Within this objective, we also set out to analyse the accuracy and precision of predictions of the time-to-contact based on the GS model. Our simulations indicated that the GS model could accurately predict the time-to-contact in various situations. We also showed that even in those contexts where it does not provide accurate estimates from motion onset, continuously updating predictions would lead to the required level of accuracy for sensorimotor tasks. Moreover, we estimated the precision of the predictions based on the GS model at a 7% of a Weber Fraction, which would indicate that the algorithm would provide reliable predictions.

The **third objective** of this thesis was to analyse if the errors estimating the time-to-contact corresponded to the predictions of the GS model. Study 2 identified different contexts where the GS model would predict the time-to-contact with systematic errors. One of those scenarios is when the trajectory is not in a collision course with the observer. Another situation would occur if the actual gravity and ball size do not correspond to the assumed a priori. Study 3 showed that the direction of the errors predicted across different trajectories corresponded to those committed by our participants for trajectories in a non-collision course. On its part, Study 4 showed that our participants committed systematic errors across gravitational accelerations and ball sizes congruent with those of the GS model. Furthermore, we show that a model based on fixed standard priors better explains our participants' temporal estimates than simple temporal mappings with optic cues or even assuming priors with the correct gravitation and ball sizes.

7.4 Limitations and future directions

The main limitation of the studies carried out during this thesis is the design of situations in a controlled experimental context. We usually recreate simplified versions of complex scenarios where relevant variables are factored out. For example, we did not include complex variables such as air resistance or the Magnus effect to simulate parabolic trajectories. We did this to create conditions as controlled and simple as possible. However, it results in trajectories that may not resemble the natural parabolic flight of an object in our day-to-day life which would be a problem to generalise our results (Aguado & López-Moliner, 2021a; Brancazio, 1985). Future studies would include these effects and examine whether the estimates of time-to-contact deviate systematically in the direction predicted by the GS model. In Study 2, we already pointed in that direction, indicating that there might be enough information to differentiate between a trajectory including air drag from another without air drag even at motion onset.

In Study 1, our participants had to estimate the direction of an object of unknown size and texture. As argued during the introductory chapter, relying only on optic cues can be problematic for extracting 3D motion information (McKee & Welch, 1989). Familiarity with the size of the object and pictorial cues would be fundamental to calibrate motion estimates in 3D coordinates (López-Moliner et al., 2007; Rushton & Duke, 2009; Zohary & Sittig, 1993). Moreover, the availability of pictorial cues has proved to improve their precision, which might reduce the bias in directional judgements (Hornsey & Hibbard, 2021).

In our experiments, binocularity provided stereoscopic information that might have helped extract the ball's absolute size (Hornsey & Hibbard, 2021; Regan & Beverley, 1979). However, this information can be incomplete or incongruent in virtual reality setups. When using an HMD or a back-projected display, there is any inconsistency between disparity and accommodation cues, which has been established to affect estimates of distance and likely motion-in-depth (Regan et al., 1986; Watt et al., 2005). Furthermore, recent studies claim that the existing HMD would not correctly adapt to the individual differences in interpupillary distance. Thus, restricting their correct use to a portion of the population or posing

generalizability problems in our results (Hibbard et al., 2020).

Furthermore, in Study 1, we extracted a measure of the variability of motion estimation by asking participants to compare lateral with motion-in-depth. Our results are congruent with those obtained by Welchman et al. (2008), in which horizontal and depth motion were estimated independently from the other. However, Manning et al. (2018b) showed that when estimating the motion of stimuli with different directions, discrimination is greatly diminished. Thus, we may have overestimated the variability in the estimation of motion-in-depth. Thus, otherwise, the standard deviation of the slow-motion prior would differ.

In Studies 2, 3 and 4, we derived our predictions from the monocular signals used by the GS model, ignoring the benefits of binocularity. Binocularity has been shown to improve the estimation of time-to-contact or performance on timed tasks. However, it is still not clear if binocularity provides radically different optic information beyond increasing the precision of our estimates (Regan, 2012). Brenner et al. (2014) argues that precision and accuracy improve because binocularity provides two independent images of the same event. Using the GS model as a basis, we can draw predictions of the resulting precision combining both independent measures. A secondary aim of Study 2 was to highlight the uncertainty in the estimation of different optic cues. It was intended to lay the groundwork for future projects where catching is modelled as a continuous problem. To this purpose, the application of tools such as Kalman filters (Kalman, 1960) could allow us to model how the precision of our estimates improves over time according to optimality criteria and compare it with human performance (Aguilar-Lleyda et al., 2018; Kwon et al., 2015).

Another limitation of the studies carried out in this thesis is the absence of an eye-tracker. Previous studies have shown that our visual tracking system is generally guided by predictions based on factors such as Earth's gravitation (Delle Monache et al., 2015; Jörges & López-Moliner, 2019) and even ball's elasticity (Diaz et al., 2013). Studying how visual tracking behaves under arbitrary gravitational accelerations in 3D parabolic trajectories is interesting to characterise the effect of incorporating a gravity prior in our sensorimotor system. Jörges and López-Moliner (2019, 2020a) shows that motion congruent with gravitational acceleration facilitates

tracking in the frontoparallel plane. However, it is unclear if those results could be directly transferable to 3D parabolic trajectories. In Study 4, we restricted the analysis of the tracking data to distinguish if the observers could track the ball effectively. Nevertheless, a quantitative analysis of tracking gains across different gravitations for 3D parabolic trajectories remains unexplored so far.

8 Conclusions

The studies presented in this thesis address how we interpret visual information. Concretely, the role of prior knowledge in interpreting 3D visual information in complex environments.

The first study explored if a slow-motion prior could explain directional biases congruently with motion estimation following Weber's law. The second study presents evidence of the use of an expectation of terrestrial gravitation and known object size. We propose the GS model as a solution based on a combination of optic and prior information to estimate the remaining time-to-contact in parabolic trajectories and devise situations where those predictions would lead to systematic errors. In the third study, we explored a context where the geometry of the presented trajectories would determine the errors predicted by the GS model and showed evidence that the errors incurred by our participants corresponded to the predicted. In the fourth study, we explored if unexpected gravitational accelerations or ball sizes lead to systematic errors as predicted by the GS model. We found that the errors committed by our participants corresponded to the errors predicted by the use of fixed priors of terrestrial gravitation and standard ball size.

Our results have significant theoretical implications regarding human rationality. Nowadays, there is an ongoing debate as to whether we use heuristic strategies to solve everyday problems (Baurès et al., 2007; Gigerenzer & Brighton, 2009; Zhao & Warren, 2015) or, on the contrary, take into account all the data available to compute a solution (de la Malla & López-Moliner, 2015; Hayhoe et al., 2005; McIntyre et al., 2001). Throughout this thesis, we provide evidence on how taking prior knowledge into account may shape the interpretation of sensory information congruently with a perspective that integrates all available information to build a final

computation.

On the other hand, it has direct practical applications. Today, commercial space expeditions are just a step away from becoming a reality. Furthermore, head-mounted displays are increasingly being used to train professional athletes, pilots, surgeons or just for fun. Virtual reality tools have the potential to shatter our expectations completely by recreating micro or supergravity environments and even non-Euclidean environments in which the size of objects is not constant. Therefore, understanding how we interpret visual information to interact with the world around us in our reality will help us develop better procedures to adapt to environments where our prior knowledge is no longer helpful.

References

Adair, R. K. (2002). *The physics of baseball*. HarperCollins, New York.

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the'light-from-above'prior. *Nature neuroscience*, *7*(10), 1057–1058.

Aguado, B., & López-Moliner, J. (2019). Perceived speed of motion in depth modulates misjudgements of approaching trajectories consistently with a slow prior. *Vision Research*, *159*, 1–9.

Aguado, B., & López-Moliner, J. (2021a). Gravity and known size calibrate visual information to time parabolic trajectories. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.642025

Aguado, B., & López-Moliner, J. (2021b). Flexible viewing time when estimating time-to-contact in 3D parabolic trajectories. *Journal of Vision*, *21*(4), 9. https://doi.org/10.1167/jov.21.4.9

Aguilar-Lleyda, D., Tubau Sala, E., & López-Moliner, J. (2018). An object-tracking model that combines position and speed explains spatial and temporal responses in a timing task. *Journal of Vision, 2018, vol. 18, num. 12, p. 1-19*.

Alam, F., Chowdhury, H., Moria, H., & Fuss, F. K. (2010). A comparative study of football aerodynamics. *Procedia Engineering*, *2*(2), 2443–2448.

Amazeen, E. L., Amazeen, P. G., Post, A. A., & Beek, P. J. (1999). Timing the selection of information during rhythmic catching. *Journal of Motor Behavior*, *31*(3), 279–289.

Anstis, S., & Kim, J. (2018). The field-size effect: Short motions look faster than long ones. *Vision research*, *146*, 32–40.

Armbrüster, C., Wolter, M., Kuhlen, T., Spijkers, W., & Fimm, B. (2008). Depth perception in virtual reality: Distance estimations in peri- and extrapersonal space. *Cyberpsychology & Behavior*, *11*(1), 9–15.

Babler, T., & Dannemiller, J. (1993). Role of image acceleration in judging landing location of free-falling projectiles. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 15–31. https://doi.org/10.1037/0096-1523.19.1.15

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Battaglia, P. W., Kersten, D., & Schrater, P. R. (2011). How haptic size sensations improve distance perception. *PLOS Computational Biology*, *7*(6), 1–13. https://doi.org/10.1371/journal.pcbi.1002080

Battaglia, P. W., Schrater, P. R., & Kersten, D. J. (2005). Auxiliary object knowledge influences visually-guided interception behavior. *Proceedings of the 2nd symposium on Applied perception in graphics and visualization*, 145–152.

Baurès, R., Benguigui, N., Amorim, M.-A., & Siegler, I. A. (2007). Intercepting free falling objects: Better use occam's razor than internalize newton's law. *Vision research*, *47*(23), 2982–2991.

Bedford, F. L. (1989). Constraints on learning new mappings between perceptual dimensions. *Journal of experimental psychology: Human perception and performance*, *15*(2), 232–248.

Belousov, B., Neumann, G., Rothkopf, C. A., & Peters, J. R. (2016). Catching heuristics are optimal control policies. *Advances in neural information processing systems*, 1426–1434.

Bennett, S. J., & Benguigui, N. (2013). Is acceleration used for ocular pursuit and spatial estimation during prediction motion? *PLoS One*, *8*(5), e63382.

Bennett, S. J., de Xivry, J.-J. O., Barnes, G. R., & Lefevre, P. (2007). Target acceleration can be extracted and represented within the predictive drive to ocular pursuit. *Journal of Neurophysiology*, *98*(3), 1405–1414.

Berkeley, G. (1709). *An essay towards a new theory of vision*. Aaron Rhames.

Berthoz, A. (2000). *The brain's sense of movement* (Vol. 10). Harvard University Press.

Beverley, K., & Regan, D. (1973). Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space. *The Journal of physiology*, *235*(1), 17–29.

Bian, Z., Braunstein, M. L., & Andersen, G. J. (2013). Discriminating direction of motion trajectories from angular speed and background information. *Attention, Perception, & Psychophysics*, *75*(7), 1570–1582.

Binaee, K., & Diaz, G. (2019). Movements of the eyes and hands are coordinated by a common predictive strategy. *Journal of Vision*, *19*(12), 3–3. https://doi.org/10.1167/19.12.3

Binaee, K., Diaz, G., Pelz, J., & Phillips, F. (2016). Binocular eye tracking calibration during a virtual ball catching task using head mounted display. *Proceedings of the acm symposium on applied perception*, 15–18.

Bingham, G. P., Bradley, A., Bailey, M., & Vinner, R. (2001). Accommodation, occlusion, and disparity matching are used to guide reaching: A comparison of actual versus virtual environments. *Journal of experimental psychology: human perception and performance*, *27*(6), 1314.

Binsted, G., Rolheiser, T., & Chua, R. (2006). Decay in visuomotor representations during manual aiming. *Journal of Motor Behavior*, *38*(2), 82–87. https://doi.org/10.3200/jmbr.38.2.82-87

Bootsma, R. J. (2009). The (current) future is here! *Perception*, *38*(6), 851–858.

Bootsma, R. J., & Oudejans, R. R. D. (1993). Visual information about time-to-collision between two objects. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(5), 1041–1052. https://doi.org/10.1037/0096-1523.19.5.1041

Bootsma, R. J., & van Wieringen, P. C. (1990). Timing an attacking forehand drive in table tennis. *Journal of experimental psychology: Human perception and performance*, *16*(1), 21–29.

Bosco, G., Delle Monache, S., & Lacquaniti, F. (2012). Catching what we can't see: Manual interception of occluded fly-ball trajectories. *PLoS One*, *7*(11), 1–20. https://doi.org/10.1371/journal.pone.0049381

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin*, *138*(3), 389.

Braddick, O., O'Brien, J. M. D., Wattam-Bell, J., Atkinson, J., Hartley, T., & Turner, R. (2001). Brain areas sensitive to coherent visual motion. *Perception*, *30*(1), 61–72. https://doi.org/10.1068/p3048

Brancazio, P. J. (1985). Looking into chapman's homer: The physics of judging a fly ball. *American Journal of Physics*, *53*(9), 849–855.

Bratzke, D., & Ulrich, R. (2021). Mental imagery of free fall: Does a falling apple accelerate in our minds? *Timing & Time Perception*, *9*(2), 150–160.

Brenner, E., Driesen, B., & Smeets, J. B. J. (2014). Precise timing when hitting falling balls. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00342

Brenner, E., & Smeets, J. B. J. (2007). Flexibility in intercepting moving objects. *Journal of Vision*, *7*(5), 1–17. https://doi.org/10.1167/7.5.14

Brenner, E., & Smeets, J. B. J. (2011). Continuous visual control of interception. *Human movement science*, *30*(3), 475–494.

Brenner, E., & Smeets, J. B. J. (2015). How people achieve their amazing temporal precision in interception. *Journal of Vision*, *15*(3), 1–21. https://doi.org/10.1167/15.3.8

Brenner, E., & Smeets, J. (2018). Continuously updating one's predictions underlies successful interception. *Journal of Neurophysiology*, *120*(6), 3257–3274. https://doi.org/10.1152/jn.00517.2018

Brenner, E., Van Den Berg, A., & Van Damme, W. (1996). Perceived motion in depth. *Vision Research*, *36*(5), 699–706. https://doi.org/https://doi.org/10.1016/0042-6989(95)00146-8

Brooks, K. R., & Stone, L. S. (2006). Stereomotion suppression and the perception of speed: Accuracy and precision as a function of 3d trajectory. *Journal of Vision*, *6*(11), 1214–1223.

Brouwer, A.-M., López-Moliner, J., Brenner, E., & Smeets, J. B. (2006). Determining whether a ball will land behind or in front of you: Not just a combination of expansion and angular velocity. *Vision Research*, *46*(3), 382–391.

Bruggeman, H., Zosh, W., & Warren, W. (2007). Optic flow drives human visuo-locomotor adaptation. *Current Biology*, *17*(23), 2035–2040. https://doi.org/10.1016/j.cub.2007.10.059

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Univ of California Press.

Burge, J., & Geisler, W. S. (2015). Optimal speed estimation in natural image movies predicts human performance. *Nature communications*, *6*(1), 1–11.

Calderone, J. B., & Kaiser, M. K. (1989). Visual acceleration detection: Effect of sign and motion orientation. *Perception & Psychophysics*, *45*(5), 391–394.

Cámara, C., de la Malla, C., López-Moliner, J., & Brenner, E. (2018). Eye movements in interception with delayed visual feedback. *Experimental brain research*, *236*(7), 1837–1847.

Chapman, S. (1968). Catching a baseball. *American Journal of Physics*, *36*(10), 868–870.

Chen, Y., Zhang, B., & Kording, K. P. (2016). Speed constancy or only slowness: What drives the kappa effect. *PLoS One*, *11*(4), e0154013.

Chénéchal, M. L., & Goldman, J. C. (2018). HTC Vive Pro Time Performance Benchmark for Scientific Research. *Icat-Egve 2018*, 81–84. https://doi.org/10.2312/egve.20181318

Chodosh, L. A., Lifson, L. E., & Tabin, C. (1995). Play ball! *Science*, *268*(5218), 1682–1683.

Craig, C. M., Berton, E., Rao, G., Fernandez, L., & Bootsma, R. J. (2006). Judging where a ball will go: the case of curved free kicks in football. *Naturwissenschaften*, *93*(2), 97–101.

Craik, K. J. W. (1967). *Nature of explanation*. Cambridge UP.

Crowell, J. A., & Banks, M. S. (1996). Ideal observer for heading judgments. *Vision Research*, *36*(3), 471–490. https://doi.org/10.1016/0042-6989(95)00121-2

Crowell, J. A., Banks, M. S., Shenoy, K. V., & Andersen, R. A. (1998). Visual self-motion perception during head turns. *Nature neuroscience*, *1*(8), 732–737.

Cumming, B. G., & Parker, A. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision research*, *34*(4), 483–495.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of space and motion* (pp. 69–117).

Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems* (Vol. 15).

de la Malla, C., & López-Moliner, J. (2015). Predictive plus online visual information optimizes temporal precision in interception. *Journal of experimental psychology: human perception and performance*, *41*(5), 1271–1280.

de la Malla, C., Smeets, J. B., & Brenner, E. (2017). Potential systematic interception errors are avoided when tracking the target with one's eyes. *Scientific reports*, *7*(1), 1–12.

de la Malla, C., Smeets, J. B., & Brenner, E. (2018). Errors in interception can be predicted from errors in perception. *Cortex*, *98*, 49–59. https://doi.org/10.1016/j.cortex.2017.03.006

de Bruyn, B., & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research*, *28*(12), 1323–1335. https://doi.org/10.1016/0042-6989(88)90064-8

Delle Monache, S., Lacquaniti, F., & Bosco, G. (2015). Eye movements and manual interception of ballistic trajectories: Effects of law of motion perturbations and occlusions. *Experimental Brain Research*, *233*(2), 359–374.

DeLucia, P. R. (2004). Multiple sources of information influence time-to-contact judgments: Do heuristics accommodate limits in sensory and cognitive processes? In *Advances in psychology* (pp. 243–285). Elsevier.

DeLucia, P. R. (2005). Does binocular disparity or familiar size information override effects of relative size on judgements of time to contact?

The Quarterly Journal of Experimental Psychology Section A, *58*(5), 865–886.

DeLucia, P. R., Meza-Arroyo, M., Baurès, R., Ranjit, M., Hsiang, S., & Gorman, J. C. (2016). Continuous response monitoring of relative time-to-contact judgments: Does effective information change during an approach event? *Ecological Psychology*, *28*(1), 1–22.

Dessing, J. C., Wijdenes, L. O., Peper, C. E., & Beek, P. J. (2009). Visuomotor transformation for interception: Catching while fixating. *Experimental Brain Research*, *196*(4), 511–527. https://doi.org/10.1007/s00221-009-1882-6

Diaz, G., Cooper, J., Rothkopf, C., & Hayhoe, M. (2013). Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of vision*, *13*(1), 1–14.

Dimitriou, M., Wolpert, D. M., & Franklin, D. W. (2013). The temporal evolution of feedback gains rapidly update to task demands. *Journal of Neuroscience*, *33*(26), 10898–10909.

Distler, H. K., Gegenfurtner, K. R., Van Veen, H. A., & Hawken, M. J. (2000). Velocity constancy in a virtual reality environment. *Perception*, *29*, 23–1435.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT press.

Elliott, D., Zuberec, S., & Milgram, P. (1994). The effects of periodic visual occlusion on ball catching. *Journal of Motor Behavior*, *26*(2), 113–122.

Ernst, M., & Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Faisal, A. A., & Wolpert, D. M. (2009). Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task. *Journal of neurophysiology*, *101*(4), 1901–1912.

Fajen, B. R. (2005a). Calibration, information, and control strategies for braking to avoid a collision. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 480–501.

Fajen, B. R. (2005b). Perceiving possibilities for action: On the necessity of calibration and perceptual learning for the visual guidance of action. *Perception*, *34*(6), 717–740.

Fajen, B. R. (2007). Affordance-based control of visually guided action. *Ecological Psychology*, *19*(4), 383–410.

Fechner, G. T. (1948/1860). Elements of psychophysics, 1860.

Fink, P. W., Foo, P. S., & Warren, W. H. (2009). Catching fly balls in virtual reality: A critical test of the outfielder problem. *Journal of vision*, *9*(13), 1–8.

Flavell, J. C. (2014). *An investigation into the directional and amplitude aspects of an internal model of gravity* (Doctoral dissertation). Manchester Metropolitan University.

Foley, J. M. (1980). Binocular distance perception. *Psychological review*, *87*(5), 411–434.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature reviews neuroscience*, *11*(2), 127–138.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*(7), 771–781. https://doi.org/10.1016/j.visres.2010.09.027

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Houghton, Mifflin; Company.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in cognitive science*, *1*(1), 107–143.

Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology*, *16*(4), 428–432.

Goldreich, D. (2007). A bayesian perceptual model replicates the cutaneous rabbit and other tactile spatiotemporal illusions. *PLoS One*, *2*(3), e333.

Gómez, J., & López-Moliner, J. (2013). Synergies between optical and physical variables in intercepting parabolic targets. *Frontiers in behavioral neuroscience*, *7*, 1–16.

Gottsdanker, R., Frick, J. W., & Lockard, R. B. (1961a). Identifying the acceleration of visual targets. *British Journal of Psychology*, *52*(1), 31–42.

Gottsdanker, R., Frick, J. W., & Lockard, R. B. (1961b). Identifying the acceleration of visual targets. *British Journal of Psychology*, *52*(1), 31–42.

Gravano, S., Zago, M., & Lacquaniti, F. (2017). Mental imagery of gravitational motion. *Cortex*, *95*, 172–191.

Gray, R. (2017). Transfer of training from virtual to real baseball batting. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02183

Gray, R., & Regan, D. (1998). Accuracy of estimating time to collision using binocular and monocular information. *Vision research*, *38*(4), 499–512.

Grealy, M. A., Craig, C. M., Bourdin, C., & Coleman, S. G. (2004). Judging time intervals using a model of perceptuo-motor control. *Journal of cognitive neuroscience*, *16*(7), 1185–1195.

Harris, D., Bird, J. M., Smart, P. A., Wilson, M. R., & Vine, S. J. (2020). A Framework for the Testing and Validation of Simulated Environments in Experimentation and Training. *Frontiers in Psychology*, *11*(March), 1–10. https://doi.org/10.3389/fpsyg.2020.00605

Harris, J., & Dean, P. J. (2003). Accuracy and precision of binocular 3-d motion perception. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 869–881.

Harris, J., & Drga, V. F. (2005). Using visual direction in three-dimensional motion perception. *Nature Neuroscience*, *8*(2), 229–233.

Harris, J., Nefs, H. T., & Grafton, C. E. (2008). Binocular vision and motion-in-depth. *Spatial vision*, *21*(6), 531–547.

Harris, J., & Watamaniuk, S. N. (1995). Speed discrimination of motion-in-depth using binocular cues. *Vision Research*, *35*(7), 885–896. https://doi.org/10.1016/0042-6989(94)00194-q

Hayhoe, M., Mennie, N., Sullivan, B., & Gorgos, K. (2005). The role of internal models and prediction in catching balls. *Proceedings of the american association for artificial intelligence*, 1–5.

Hecht, H., Kaiser, M. K., & Banks, M. S. (1996). Gravitational acceleration as a cue for absolute size and distance? *Perception & Psychophysics*, *58*(7), 1066–1075.

Hecht, H., & Savelsbergh, G. (2004). *Time-to-contact* (Vol. 135). Elsevier.

Helmholtz, H. V. (1867). *Treatise on physiological optics vol. iii*.

Heuer, H. (1993). Estimates of time to contact based on changing size and changing target vergence. *Perception*, *22*(5), 549–563. https://doi.org/10.1068/p220549

Hibbard, P. B., van Dam, L. C., & Scarfe, P. (2020). The implications of interpupillary distance variability for virtual reality. *2020 International Conference on 3D Immersion (IC3D)*, 1–7.

Höfer, S., Raisch, J., Toussaint, M., & Brock, O. (2018). No free lunch in ball catching: A comparison of Cartesian and angular representations for control. *PLoS One*, *13*(6), 1–48. https://doi.org/10.1371/journal.pone.0197803

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, *8*(3), 33–33.

Hornsey, R. L., & Hibbard, P. B. (2021). Contributions of pictorial and binocular cues to the perception of distance in virtual reality. *Virtual Reality*, *25*(4), 1087–1103.

Hosking, S. G., & Crassini, B. (2010). The effects of familiar size and object trajectories on time-to-contact judgements. *Experimental Brain Research*, *203*(3), 541–552. https://doi.org/10.1007/s00221-010-2258-7

Hu, Y., & Goodale, M. (2000). Grasping after a delay shifts size-scaling from absolute to relative metrics. *Journal of Cognitive Neuroscience*, *12*(5), 856–868.

Hubbard, T. L. (1995). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic bulletin & review*, *2*(3), 322–338.

Indovina, I. (2005). Representation of Visual Gravitational Motion in the Human Vestibular Cortex. *Science*, *308*(5720), 416–419. https://doi.org/10.1126/science.1107961

Ittelson, W. H. (1951). Size as a cue to distance: static localization. *The American journal of psychology*, *64*(1), 54–67. https://doi.org/10.2307/1418595

Jacobs, D. M., & Michaels, C. F. (2006). Lateral interception I: Operative optical variables, attunement, and calibration. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 443–458. https://doi.org/10.1037/0096-1523.32.2.443

Jacobs, D. M., & Michaels, C. F. (2007). Direct learning. *Ecological psychology*, *19*(4), 321–349.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature neuroscience*, *13*(8), 1020–1026.

Joerges, B., Hagenfeld, L., & Lopez-Moliner, J. (2018). The use of visual cues in gravity judgements on parabolic motion. *Vision research*, *149*, 47–58.

Jokisch, D., & Troje, N. F. (2003). Biological motion as a cue for the perception of size. *Journal of Vision*, *3*(4), 1–1.

Jörges, B., La Scaleia, B., López-Moliner, J., Lacquaniti, F., & Zago, M. (2021). Perceptual judgments of duration of parabolic motions. *Scientific reports*, *11*(1), 1–13.

Jörges, B., & López-Moliner, J. (2017). Gravity as a strong prior: Implications for perception and action. *Frontiers in Human Neuroscience*, *11*, 47–58. https://doi.org/10.3389/fnhum.2017.00203

Jörges, B., & López-Moliner, J. (2019). Earth-gravity congruent motion facilitates ocular control for pursuit of parabolic trajectories. *Scientific Reports*, *9*(1), 1–13.

Jörges, B., & López-Moliner, J. (2020a). Determining mean and standard deviation of the strong gravity prior through simulations. *PLoS One*, *15*(8), e0236732.

Jörges, B., & López-Moliner, J. (2020b). Determining mean and standard deviation of the strong gravity prior through simulations. *PLoS One*, *15*(8), 1–19. https://doi.org/10.1371/journal.pone.0236732

Judge, S., & Rind, F. (1997). The locust dcmd, a movement-detecting neurone tightly tuned to collision trajectories. *Journal of Experimental Biology*, *200*(16), 2209–2216.

Kagan, D., & Nathan, A. M. (2014). Simplified models for the drag coefficient of a pitched baseball. *The Physics Teacher*, *52*(5), 278–280.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552

Katsumata, H., & Russell, D. M. (2012). Prospective versus predictive control in timing of hitting a falling ball. *Experimental Brain Research*, *216*(4), 499–514. https://doi.org/10.1007/s00221-011-2954-y

Keele, S. W., & Posner, M. I. (1968). Processing of Visual Feedback in Rapid Movements. *Journal of Experimental Psychology*, *77*(1), 155–158. https://doi.org/10.1037/h0025754

Keil, M. S., & López-Moliner, J. (2012). Unifying time to contact estimation and collision avoidance across species. *PLoS computational biology*, *8*(8), e1002625.

Kelly, J. W., Donaldson, L. S., Sjolund, L. A., & Freiberg, J. B. (2013). More than just perception–action recalibration: Walking through a virtual environment causes rescaling of perceived space. *Attention, Perception, & Psychophysics*, *75*(7), 1473–1485.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, *55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Khoei, M. A., Masson, G. S., & Perrinet, L. U. (2017). The flash-lag effect as a motion-based predictive shift. *PLoS computational biology*, *13*(1), e1005068.

Kistemaker, D. A., Faber, H., & Beek, P. J. (2009). Catching fly balls: A simulation study of the Chapman strategy. *Human Movement Science*, *28*(2), 236–249. https://doi.org/10.1016/j.humov.2008.11.001

Kistemaker, D. A., Van Soest, A. (J., & Bobbert, M. F. (2006). Is equilibrium point control feasible for fast goal-directed single-joint movements? *Journal of Neurophysiology*, *95*(5), 2898–2912.

Klein, S. A., & Levi, D. M. (1987). Position sense of the peripheral retina. *JOSA A*, *4*(8), 1543–1553.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS one*, *2*(e943), 43.

Kwon, O. S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(26), 8142–8147. https://doi.org/10.1073/pnas.1500361112

Lacquaniti, F., & Maioli, C. (1989). The role of preparation in tuning anticipatory and reflex responses during catching. *The Journal of Neuroscience*, *9*(1), 134–148. https://doi.org/10.1523/jneurosci.09-01-00134.1989

Lages, M. (2006). Bayesian models of binocular 3-d motion perception. *Journal of Vision*, *6*(4), 14–14.

Lages, M., & Heron, S. (2010). On the inverse problem of binocular 3d motion perception. *PLoS computational biology*, *6*(11), e1000999.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception*, *5*(4), 437–459.

Lee, D. N., Young, D., Reddish, P., Lough, S., & Clayton, T. (1983). Visual timing in hitting an accelerating ball. *The Quarterly Journal of Experimental Psychology*, *35*(2), 333–346.

Lewis, R. F., Gaymard, B. M., & Tamargo, R. J. (1998). Efference copy provides the eye position information required for visually guided reaching. *Journal of Neurophysiology*, *80*(3), 1605–1608. https://doi.org/10.1152/jn.1998.80.3.1605

Linares, D., & López-Moliner, J. (2016). Quickpsy: An r package to fit psychometric functions for multiple groups. *The R Journal, 2016, vol. 8, num. 1, p. 122-131*.

Liu, D., & Todorov, E. (2007). Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *Journal of Neuroscience*, *27*(35), 9354–9368.

López-Moliner, J., & Brenner, E. (2016). Flexible timing of eye movements when catching a ball. *Journal of Vision*, *16*(5), 1–11. https://doi.org/10.1167/16.5.13

López-Moliner, J., Brenner, E., Louw, S., & Smeets, J. B. (2010). Catching a gently thrown ball. *Experimental Brain Research*, *206*(4), 409–417. https://doi.org/10.1007/s00221-010-2421-1

López-Moliner, J., Field, D. T., & Wann, J. P. (2007). Interceptive timing: Prior knowledge matters. *Journal of Vision*, *7*(13), 1–8.

López-Moliner, J., & Keil, M. S. (2012). People Favour Imperfect Catching by Assuming a Stable World (M. O. Ernst, Ed.). *PLoS One*, *7*(4), 1–8. https://doi.org/10.1371/journal.pone.0035705

López-Moliner, J., Supèr, H., & Keil, M. S. (2013). The time course of estimating time-to-contact: Switching between sources of information. *Vision research*, *92*, 53–58.

López-Moliner, J., Vullings, C., Madelain, L., & Van Beers, R. J. (2019). Prediction and final temporal errors are used for trial-to-trial motor corrections. *Scientific Reports*, *9*(1), 1–15.

Maloney, L. T., & Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research*, *50*(23), 2362–2374. https://doi.org/10.1016/j.visres.2010.09.031

Maltz, M. V., Stubbs, K. M., Quinlan, D. J., Rzepka, A. M., Martin, J. R., & Culham, J. C. (2021). Familiar size affects the perceived size and distance of real objects even with binocular vision. *Journal of vision*, *21*(10), 21–21.

Mann, D. L., Spratford, W., & Abernethy, B. (2013). The Head Tracks and Gaze Predicts: How the World's Best Batters Hit a Ball. *PLoS One*, *8*(3), 1–11. https://doi.org/10.1371/journal.pone.0058289

Manning, C., Thomas, R. T., & Braddick, O. (2018a). Can speed be judged independent of direction? *Journal of Vision*, *18*(6), 15–15.

Manning, C., Thomas, R. T., & Braddick, O. (2018b). Can speed be judged independent of direction? *Journal of vision*, *18*(6), 15–15.

Manuilova, E., Schuetzenmeister, A., & Model, F. (2021). Package 'mcr'.

Mazyn, L. I., Savelsbergh, G. J., Montagne, G., & Lenoir, M. (2007). Planning and on-line control of catching as a function of perceptual-

motor constraints. *Acta Psychologica*, *126*(1), 59–78. https://doi.org/10.1016/j.actpsy.2006.10.001

McBeath, M. K., Nathan, A. M., Bahill, A. T., & Baldwin, D. G. (2008). Paradoxical pop-ups: Why are they difficult to catch? *American Journal of Physics*, *76*(8), 723–729.

McBeath, M. K., Shaffer, D., & Kaiser, M. (1995). How baseball outfielders determine where to run to catch fly balls. *Science*, *268*(5210), 569–573. https://doi.org/10.1126/science.7725104

McBeath, M. K., Tang, T. Y., & Shaffer, D. M. (2018). The geometry of consciousness. *Consciousness and cognition*, *64*, 207–215. https://doi.org/10.1016/j.concog.2018.04.015

McConnell, D. S., Muchisky, M. M., & Bingham, G. P. (1998). The use of time and trajectory forms as visual information about spatial scale in events. *Perception & Psychophysics*, *7*(60), 1175–1187.

McIntyre, J., Senot, P., Prevost, P., Zago, M., Lacquaniti, F., & Berthoz, A. (2003). The use of on-line perceptual invariants versus cognitive internal models for the predictive control of movement and action. *First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings.*, 438–441.

McIntyre, J., Zago, M., Berthoz, A., & Lacquaniti, F. (2001). Does the brain model newton's laws? *Nature neuroscience*, *4*(7), 693–694.

McKee, S. P. (1981). A local mechanism for differential velocity detection. *Vision research*, *21*(4), 491–500.

McKee, S. P., Silverman, G. H., & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision research*, *26*(4), 609–619.

McKee, S. P., & Welch, L. (1989). Is there a constancy for velocity? *Vision research*, *29*(5), 553–561.

McKee, S. P., & Welch, L. (1992). The precision of size constancy. *Vision Research*, *32*(8), 1447–1460. https://doi.org/10.1016/0042-6989(92)90201-S

McLeod, P., & Dienes, Z. (1993). Running to catch the ball. *Nature*, *362*(6415), 23–23. https://doi.org/10.1038/362023a0

McLeod, P., & Dienes, Z. (1996). Do fielders know where to go to catch the ball or only how to get there? *Journal of experimental psychology: human perception and performance*, *22*(3), 531–543.

McLeod, P., Reed, N., & Dienes, Z. (2006). The generalized optic acceleration cancellation theory of catching. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 139–148.

Miall, R., Weir, D., & Stein, J. (1986). Manual tracking of visual targets by trained monkeys. *Behavioural brain research*, *20*(2), 185–201.

Michaels, C. F., & Oudejans, R. R. (1992). The Optics and Actions of Catching Fly Balls: Zeroing Out Optical Acceleration. *Ecological Psychology*, *4*(4), 199–222. https://doi.org/10.1207/s15326969eco0404_1

Miller, W. L., Maffei, V., Bosco, G., Iosa, M., Zago, M., Macaluso, E., & Lacquaniti, F. (2008). Vestibular nuclei and cerebellum put visual gravitational motion in context. *Journal of neurophysiology*, *99*(4), 1969–1982.

Montagne, G., Laurent, M., Durey, A., & Bootsma, R. (1999). Movement reversals in ball catching. *Experimental Brain Research*, *129*(1), 87–92.

Moran, A., Campbell, M., & Ranieri, D. (2018). Implications of eye tracking technology for applied sport psychology. *Journal of Sport Psychology in Action*, *9*(4), 249–259. https://doi.org/10.1080/21520704.2018.1511660

Moscatelli, A., & Lacquaniti, F. (2011). The weight of time: Gravitational force enhances discrimination of visual motion duration. *Journal of Vision*, *11*(4), 1–17. https://doi.org/10.1167/11.4.5

Moscatelli, A., La Scaleia, B., Zago, M., & Lacquaniti, F. (2019). Motion direction, luminance contrast, and speed perception: An unexpected meeting. *Journal of Vision*, *19*(6), 16–16.

Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2005). Neuronal computation of disparity in v1 limits temporal resolution for detecting disparity modulation. *Journal of Neuroscience*, *25*(44), 10207–10219.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, *370*(6487), 256–257.

Oberle, C., McBeath, M. K., Madigan, S., & Sugar, T. (2005). The galileo bias: A naive conceptual belief that influences people's perceptions and performance in a ball-dropping task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 643–653. https://doi.org/10.1037/0278-7393.31.4.643

Orban, G. A., de Wolf, J., & Maes, H. (1984). Factors influencing velocity coding in the human visual system. *Vision Research*, *24*(1), 33–39. https://doi.org/10.1016/0042-6989(84)90141-X

Oudejans, R. R. D., Michaels, C. F., & Bakker, F. C. (1997). The effects of baseball experience on movement initiation in catching fly balls. *Journal of Sports Sciences*, *15*(6), 587–595. https://doi.org/10.1080/026404197367029

Oudejans, R. R. D., Michaels, C. F., Bakker, F. C., & Davids, K. (1999). Shedding some light on catching in the dark: Perceptual mechanisms for catching fly balls. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 531–542.

Peper, L., Bootsma, R. J., Mestre, D. R., & Bakker, F. C. (1994). Catching Balls: How to Get the Hand to the Right Place at the Right Time. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(3), 591–612. https://doi.org/10.1037/0096-1523.20.3.591

Perrone, J. A. (2006). A single mechanism can explain the speed tuning properties of mt and v1 complex neurons. *Journal of Neuroscience*, *26*(46), 11987–11991.

Peters, M. A. K., Ma, W. J., & Shams, L. (2016). The Size-Weight Illusion is not anti-Bayesian after all: a unifying Bayesian account. *PeerJ*, *4*, e2124.

Peters, M. A., Zhang, L.-Q., & Shams, L. (2018). The material-weight illusion is a bayes-optimal percept under competing density priors. *PeerJ*, *6*, e760.

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*(24), 3145–3161. https://doi.org/10.1016/S0042-6989(01)00173-0

Poljac, E., Neggers, B., & Van Den Berg, A. (2006). Collision judgment of objects approaching the head. *Experimental brain research*, *171*(1), 35–46.

Ponce, C. R., Lomber, S. G., & Born, R. T. (2008). Integrating motion and depth via parallel pathways. *Nature neuroscience*, *11*(2), 216.

Portfors-Yeomans, C. V., & Regan, D. (1996). Cyclopean discrimination thresholds for the direction and speed of motion in depth. *Vision research*, *36*(20), 3265–3279. https://doi.org/10.1016/0042-6989(96)00065-X

Postma, D. B. W., Smith, J., Pepping, G.-J., van Andel, S., & Zaal, F. T. J. M. (2017). When a Fly Ball Is Out of Reach: Catchability Judgments Are Not Based on Optical Acceleration Cancelation. *Frontiers in Psychology*, *8*, 535. https://doi.org/10.3389/fpsyg.2017.00535

Postma, D. B., Den Otter, A. R., & Zaal, F. T. (2014). Keeping your eyes continuously on the ball while running for catchable and uncatchable fly balls. *PLoS One*, *9*(3), 1–5. https://doi.org/10.1371/journal.pone.0092392

Postma, D. B., Lemmink, K. A., & Zaal, F. T. (2018). The affordance of catchability in running to intercept fly balls. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(9), 1336–1347.

Qian, J., & Yazdanbakhsh, A. (2015). A neural model of distance-dependent percept of object size constancy. *PLoS One*, *10*, 129377.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rahnev, D., & Denison, R. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*(e223), 1–66.

Reed, N., McLeod, P., & Dienes, Z. (2010). Implicit knowledge and motor skill: What people who know how to catch don't know. *Consciousness and cognition*, *19*(1), 63–76.

Regan, D., Erkelens, C. J., & Collewijn, H. (1986). Necessary conditions for the perception of motion in depth. *Investigative Ophthalmology & Visual Science*, *27*(4), 584–597.

Regan, D., & Hamstra, S. (1993). Dissociation of discrimination thresholds for time to contact and for rate of angular expansion. *Vision Research*, *33*(4), 447–462. https://doi.org/10.1016/0042-6989(93)90252-R

Regan, D. (1992). Visual judgements and misjudgements in cricket, and the art of flight. *Perception*, *21*(1), 91–115.

Regan, D. (2012). Vision and cricket. *Ophthalmic and Physiological Optics*, *32*(4), 257–270.

Regan, D., & Beverley, K. I. (1979). Binocular and monocular stimuli for motion in depth: Changing-disparity and changing-size feed the same motion-in-depth stage. *Vision research*, *19*(12), 1331–1342.

Regan, D., & Beverley, K. (1978). Looming detectors in the human visual pathway. *Vision research*, *18*(4), 415–421.

Regan, D., & Kaushal, S. (1994). Monocular discrimination of the direction of motion in depth. *Vision Research*, *34*(2), 163–177.

Rind, F. C., & Simmons, P. J. (1999). Seeing what is coming: Building collision-sensitive neurones. *Trends in neurosciences*, *22*(5), 215–220.

Rock, I., Hill, A. L., & Fineman, M. (1968). *Speed constancy as a function of size constancy*.

Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity-and velocity-based signals for three-dimensional motion perception in human mt+. *Nature neuroscience*, *12*(8), 1050.

Rokers, B., Fulvio, J. M., Pillow, J. W., & Cooper, E. A. (2018). Systematic misperceptions of 3-d motion explained by bayesian inference. *Journal of vision*, *18*(3), 23–23.

Rolin, R. A., Fooken, J., Spering, M., & Pai, D. K. (2018). Perception of looming motion in virtual reality egocentric interception tasks. *IEEE transactions on visualization and computer graphics*, *25*(10), 3042–3048.

Roy, J. E., & Cullen, K. E. (2003). Brain stem pursuit pathways: Dissociating visual, vestibular, and proprioceptive inputs during combined eye-head gaze tracking. *Journal of neurophysiology*, *90*(1), 271–290.

RStudio Team. (2020). *Rstudio: Integrated development environment for r*. RStudio, PBC. Boston, MA. http://www.rstudio.com/

Rushton, S. K. (2004). Interception of projectiles, from when & where to where once. In *Advances in psychology* (pp. 327–353). Elsevier.

Rushton, S. K., & Duke, P. A. (2009). Observers cannot accurately estimate the speed of an approaching object in flight. *Vision research*, *49*(15), 1919–1928.

Rushton, S. K., & Wann, J. P. (1999). Weighted combination of size and disparity: A computational model for timing a ball catch. *Nature Neuroscience*, *2*(2), 186–190. https://doi.org/10.1038/5750

Russo, M., Cesqui, B., La Scaleia, B., Ceccarelli, F., Maselli, A., Moscatelli, A., Zago, M., Lacquaniti, F., & D'Avella, A. (2017). Intercepting virtual balls approaching under different gravity conditions: evidence for spatial prediction. *Journal of Neurophysiology*, *118*(4), 2421–2434. https://doi.org/10.1152/jn.00025.2017

Samad, M., Chung, A. J., & Shams, L. (2015). *Perception of body ownership is driven by bayesian sensory inference*.

Sanada, T. M., & DeAngelis, G. C. (2014). Neural representation of motion-in-depth in area mt. *Journal of Neuroscience*, *34*(47), 15508–15521.

Savelsbergh, G., & Whiting, H. (1992). The acquisition of catching under monocular and binocular conditions. *Journal of Motor Behavior*, *24*(4), 320–328. https://doi.org/10.1080/00222895.1992.9941628

Savelsbergh, G., Whiting, H., & Bootsma, R. J. (1991). Grasping tau. *Journal of experimental psychology: human perception and performance*, *17*(2), 315.

Saxberg, B. (1987a). Projected free fall trajectories. i. theory and simulation. *Biological cybernetics*, *56*(2-3), 159–176.

Saxberg, B. (1987b). Projected free fall trajectories. ii. human experiments. *Biological Cybernetics*, *56*(2-3), 177–184.

Schütz, A. C., Braun, D. I., Kerzel, D., & Gegenfurtner, K. R. (2008). Improved visual sensitivity during smooth pursuit eye movements. *Nature neuroscience*, *11*(10), 1211.

Series, P., Georges, S., Lorenceau, J., & Frégnac, Y. (2002). Orientation dependent modulation of apparent speed: A model based on the

dynamics of feed-forward and horizontal connectivity in v1 cortex. *Vision research*, *42*(25), 2781–2797.

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual review of neuroscience*, *33*, 89–108.

Shaffer, D. M., Marken, R. S., Dolgov, I., & Maynor, A. B. (2013). Chasin'choppers: Using unpredictable trajectories to test theories of object interception. *Attention, Perception, & Psychophysics*, *75*(7), 1496–1506.

Shaffer, D. M., & McBeath, M. K. (2005). Naive beliefs in baseball: systematic distortion in perceived time of apex for fly balls. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1492–1501. https://doi.org/10.1037/0278-7393.31.6.1492

Sharp, R. H., & Whiting, H. T. (1974). Exposure and occluded duration effects in a ball-catching skill. *Journal of Motor Behavior*, *6*(3), 139–147. https://doi.org/10.1080/00222895.1974.10734990

Smith, M. R., Flach, J. M., Dittman, S. M., & Stanard, T. (2001). Monocular optical constraints on collision control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(2), 395–410. https://doi.org/10.1037/0096-1523.27.2.395

Snowden, R. J., & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision research*, *31*(5), 907–914.

Soechting, J. F., Juveli, J. Z., & Rao, H. M. (2009). Models for the extrapolation of target motion for manual interception. *Journal of neurophysiology*, *102*(3), 1491–1502.

Sousa, R., Brenner, E., & Smeets, J. B. J. (2011). Judging an unfamiliar object ' s distance from its retinal image size. *Journal of vision*, *11*(9), 1–6. https://doi.org/10.1167/11.9.10.Introduction

Spering, M., Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Keep your eyes on the ball: smooth pursuit eye movements enhance prediction of visual motion. *Journal of Neurophysiology*, *105*(4), 1756–1767.

Stocker, A. A., & Simoncelli, E. P. (2005). Constraining the prior and likelihood in a bayesian model of human visual speed perception. *Journal of Vision*, *5*(8), 928.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, *9*(4), 578–585.

Sun, H., & Frost, B. (1998). Computation of different optical variables of looming objects in pigeon nucleus rotundus neurons. *Nature Neuroscience*, *1*(4), 296–303. https://doi.org/10.1038/1110

Tcheang, L., Gilson, S. J., & Glennerster, A. (2005). Systematic distortions of perceptual stability investigated using immersive virtual reality. *Vision Research*, *45*(16), 2177–2189. https://doi.org/10.1016/j.visres.2005.02.006

Thomas, O. M., Cumming, B. G., & Parker, A. J. (2002). A specialization for relative disparity in v2. *Nature neuroscience*, *5*(5), 472.

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision research*, *22*, 7–380.

Thompson, P., Brooks, K., & Hammett, S. T. (2006). Speed can go up as well as down at low contrast: Implications for models of motion perception. *Vision research*, *46*(6-7), 782–786.

Timmerman, P., & van der Weele, J. (1999). On the rise and fall of a ball with linear or quadratic drag. *American Journal of Physics*, *67*(6), 538–546. https://doi.org/10.1119/1.19320

Todd, J. T. (1981). Visual information about moving objects. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 795–810. https://doi.org/10.1037/0096-1523.7.4.795

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, *7*(9), 907–915. https://doi.org/10.1038/nn1309

Tresilian, J. R. (1994). Approximate information sources and perceptual variables in interceptive timing. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 154–173. https://doi.org/10.1037/0096-1523.20.1.154

Tresilian, J. R. (1995). Study of a Servo-control Strategy for Projectile Interception. *The Quarterly Journal of Experimental Psychology Section A*, *48*(3), 688–715. https://doi.org/10.1080/14640749508401411

Trommershauser, J., Maloney, L. T., & Landy, M. S. (2003). Statistical decision theory and trade-offs in the control of motor response. *Spatial vision*, *16*(3), 255–275.

Turvey, M., Shaw, R., Reed, E., & Mace, W. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition*, *9*(3), 237–304. https://doi.org/10.1016/0010-0277(81)90002-0

Vagnoni, E., Lourenco, S. F., & Longo, M. R. (2015). Threat modulates neural responses to looming visual stimuli. *European Journal of Neuroscience*, *42*(5), 2190–2202.

van Dam, L. C., & Ernst, M. O. (2015). Mapping shape to visuomotor mapping: Learning and generalisation of sensorimotor behaviour based on contextual information. *PLoS computational biology*, *11*(3), e1004172.

Vintch, B., & Gardner, J. L. (2014). Cortical correlates of human motion perception biases. *Journal of Neuroscience*, *34*(7), 2592–2604.

Wann, J. P. (1996). Anticipating Arrival: Is the Tau Margin a Specious Theory? *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 1031–1048. https://doi.org/10.1037/0096-1523.22.4.1031

Warren, W., Kay, B., Zosh, W., Duchon, A., & Sahuc, S. (2001). Optic flow is used to control human walking. *Nature Neuroscience*, *4*(2), 213–216. https://doi.org/10.1038/84054

Watson, A. B., & Pelli, D. G. (1983). Quest: A bayesian adaptive psychometric method. *Perception & psychophysics*, *33*(2), 113–120.

Watson, J., Banks, M. S., von Hofsten, C., & Royden, C. S. (1992). Gravity as a monocular cue for perception of absolute distance and/or absolute size. *Perception*, *21*(1), 69–76.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of vision*, *5*(10), 7–7.

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience*, *18*(10), 1509–1517. https://doi.org/10.1038/nn.4105

Wei, X.-X., & Stocker, A. A. (2017). Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences*, *114*(38), 10244–10249.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature neuroscience*, *5*(6), 598–604.

Welch, L. (1989). The perception of moving plaids reveals two motion-processing stages. *Nature*, *337*(6209), 734.

Welchman, A. E., Lam, J. M., & Bülthoff, H. H. (2008). Bayesian motion estimation accounts for a surprising bias in 3d vision. *Proceedings of the National Academy of Sciences*, *105*(33), 12087–12092.

Welchman, A. E., Tuck, V. L., & Harris, J. (2004). Human observers are biased in judging the angular approach of a projectile. *Vision Research*, *44*(17), 2027–2042.

Werkhoven, P., Snippe, H. P., & Alexander, T. (1992). Visual processing of optic acceleration. *Vision research*, *32*(12), 2313–2329.

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision research*, *17*(8), 941–947.

Whiting, H. T. (1968). Training in a Continuous Ball Throwing and Catching Task. *Ergonomics*, *11*(4), 375–382. https://doi.org/10.1080/00140136808930985

Whiting, H., & Sharp, R. (1974). Visual occlusion factors in a discrete ball-catching task. *Journal of Motor Behavior*, *6*(1), 11–16. https://doi.org/10.1080/00222895.1974.10734974

Wilkie, R., & Wann, J. (2003). Controlling steering and judging heading: Retinal flow, visual direction, and extraretinal information. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 363–378. https://doi.org/10.1037/0096-1523.29.2.363

Willemsen, P., Gooch, A. A., Thompson, W. B., & Creem-Regehr, S. H. (2008). Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments*, *17*(1), 91–101.

Wilson, A. D., Golonka, S., & Barrett, L. (2013). Embodied Cognition is Not What you Think it is. *Frontiers in Psychology*, *4*(2), 1–13. https://doi.org/10.3389/fpsyg.2013.00058

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Yonas, A., Bechtold, A. G., Frankel, D., Gordon, F. R., McRoberts, G., Norcia, A., & Sternfels, S. (1977). Development of sensitivity to information for impending collision. *Perception & Psychophysics*, *21*(2), 97–104.

Zaal, F., Bongers, R., Pepping, G., & Bootsma, R. (2012). Base on balls for the chapman strategy: Reassessing brouwer, brenner, and smeets (2002). *Attention, Perception, & Psychophysics*, *74*(7), 1488–1498.

Zaal, F. T. J. M., & Bootsma, R. J. (2011). Virtual Reality as a Tool for the Study of Perception-Action: The Case of Running to Catch Fly Balls. *Presence: Teleoperators and Virtual Environments*, *20*(1), 93–103. https://doi.org/10.1162/pres_a_00037

Zago, M., Bosco, G., Maffei, V., Iosa, M., Ivanenko, Y. P., & Lacquaniti, F. (2004). Internal models of target motion: expected dynamics overrides measured kinematics in timing manual interceptions. *Journal of neurophysiology*, *91*(4), 1620–1634.

Zago, M., Bosco, G., Maffei, V., Iosa, M., Ivanenko, Y., & Lacquaniti, F. (2005). Fast adaptation of the internal model of gravity for manual interceptions: Evidence for event-dependent learning. *Journal of Neurophysiology*, *93*(2), 1055–1068. https://doi.org/10.1152/jn.00833.2004

Zago, M., & Lacquaniti, F. (2005). Internal model of gravity for hand interception: Parametric adaptation to zero-gravity visual targets on earth. *Journal of Neurophysiology*, *94*(2), 1346–1357. https://doi.org/10.1152/jn.00215.2005

Zago, M., McIntyre, J., Senot, P., & Lacquaniti, F. (2008). Internal models and prediction of visual gravitational motion. *Vision Research*, *48*(14), 1532–1538. https://doi.org/10.1016/J.VISRES.2008.04.005

Zago, M., McIntyre, J., Senot, P., & Lacquaniti, F. (2009). Visuo-motor coordination and internal models for object interception. *Experimental Brain Research*, *192*(4), 571–604.

Zhao, H., Straub, D., & Rothkopf, C. A. (2019). The visual control of interceptive steering: How do people steer a car to intercept a moving target? *Journal of Vision*, *19*(14), 1–20.

Zhao, H., & Warren, W. H. (2015). On-line and model-based approaches to the visual control of action. *Vision research*, *110*, 190–202. https://doi.org/10.1016/j.visres.2014.10.008

Zohary, E., & Sittig, A. C. (1993). Mechanisms of velocity constancy. *Vision research*, *33*(17), 2467–2478.

List of abreviations

- HMD: Head-Mounted Display
- TTC: Time-to-contact / Time-to-collision
- MID: Mition-in-depth
- LMM: Linear-Mixed Model
- V_x: Horizontal displacement
- V_y: Vertical displacement
- V_z: Depth displacement
- G: Gravitational acceleration
- s: Object's size

List of Figures

List of Tables

Front pages

Study 1

Vision Research 159 (2019) 1–9

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Perceived speed of motion in depth modulates misjudgements of approaching trajectories consistently with a slow prior

Borja Aguado, Joan López-Moliner*

Vision and Control of Action (VISCA) Group, Department of Cognition, Development and Psychology of Education, Institut de Neurociències, Universitat de Barcelona, Passeig de la Vall d'Hebron 171, 08035 Barcelona, Catalonia, Spain

ARTICLE INFO

Keywords:
3D perception
Bayesian models
Motion-in-depth
Speed perception
Virtual reality
Weber's law

ABSTRACT

Previous studies have shown that the angle of approach is consistently overestimated for approaching (but passing-by) objects. An explanation based on a slow-motion prior has been proposed in the past to account for this bias. The mechanism relies on the (less reliable) in-depth component of the motion being more attracted towards the slow motion prior than the (more reliable) lateral component. This hypothesis predicts that faster speeds in depth will translate into a greater bias if the perception of velocity in depth follows Weber's law. Our approach is different than the one used in previous studies where perceived speed and direction were measured in different experiments. To test our hypothesis, we conducted an experiment in which participants estimated approaching angles via a pointing device, while at the same time comparing the speed of the approaching object with a lateral velocity reference. This way, we couple perceived speed with perceived trajectory for each approaching angle in the same trial. Our results show that the directional bias is larger for faster objects, which is consistent with motion in depth following Weber's law. The differential biases can be accounted for by a Bayesian model that includes a slow motion prior.

1. Introduction

One of the main functions of the visual system is to recover the 3D structure of the environment. This is particularly important when we need to estimate the direction and speed of moving objects on a collision (or near-collision) course with us.

Knowing how different cues, both monocular and binocular, contribute to estimate direction and motion in depth (MID) has attracted interest in the past (Beverley & Regan, 1973; Cumming & Bruce, 1994), but still is an active field of research (Harris, Nefs, & Grafton, 2008; Rokers, Fulvio, Pillow, & Cooper, 2018). Past work on MID, however, has mainly focused on precision and accuracy of motion estimates (Harris & Dean, 2003; Rushton & Duke, 2009).

Regarding the direction of approach, several studies have shown that we tend to overestimate the bearing angle (from now on β; see Fig. 1) of the trajectory of an approaching target (Harris & Drga, 2005; Lages, 2006; Poljac, Neggers, & Van Den Berg, 2006; Welchman, Tuck, & Harris, 2004). This is, we overestimate the lateral distance by which a ball passes us. This can be counterintuitive, given that we are very sensitive to the motion direction of objects on a collision course (Regan, Erkelens, & Collewijn, 1986).

To explain this bias, Welchman, Lam, and Bülthoff (2008) put forward a Bayesian explanation that included the so-called *Slow Motion Prior*, which is a main component of a motion perception model by Stocker and Simoncelli (2006). Sensory estimates (likelihood) are combined with an expectation of nearly zero motion in the environment (prior) resulting in consistent underestimations of speed (posterior), with the extent of underestimation depending on the reliability of the likelihood (e.g. contrast of a grating; Stocker & Simoncelli (2006)). Therefore, if the reliability of the signal is low, the slow prior will be weighted more, resulting in a slower posterior and, consequently, the speed of the stimulus will be underestimated more strongly.

In the same study, Stocker and Simoncelli (2006) found that the width of likelihood estimates for speed discrimination tasks increases logarithmically as a function of speed approximately following Weber's law (for targets moving faster than 1 deg/s), as suggested by previous literature in the field (McKee, Silverman, & Nakayama, 1986; Welch, 1989). Furthermore, they used a Bayesian Observer model to infer the shape (SD) of the Slow Motion Prior, which falls from a peak at slow speeds becoming shallower for faster ones. As a result, the prior expectation introduces increasingly biases for the posterior as a function of the perceived speed.

Welchman et al. (2008) explained the underestimation of approaching angles in terms of this slow prior: The estimate of the lateral

* Corresponding author.
E-mail address: j.lopezmoliner@ub.edu (J. López-Moliner).

https://doi.org/10.1016/j.visres.2019.03.009
Received 28 September 2018; Received in revised form 12 March 2019; Accepted 12 March 2019
Available online 28 March 2019

Study 2

frontiers in Human Neuroscience

HYPOTHESIS AND THEORY
published: 23 August 2021
doi: 10.3389/fnhum.2021.642025

Gravity and Known Size Calibrate Visual Information to Time Parabolic Trajectories

*Borja Aguado and Joan López-Moliner**

Vision and Control of Action (VISCA) Group, Department of Cognition, Development and Psychology of Education, Institut de Neurociències, Universitat de Barcelona, Barcelona, Spain

Catching a ball in a parabolic flight is a complex task in which the time and area of interception are strongly coupled, making interception possible for a short period. Although this makes the estimation of time-to-contact (TTC) from visual information in parabolic trajectories very useful, previous attempts to explain our precision in interceptive tasks circumvent the need to estimate TTC to guide our action. Obtaining TTC from optical variables alone in parabolic trajectories would imply very complex transformations from 2D retinal images to a 3D layout. We propose based on previous work and show by using simulations that exploiting prior distributions of gravity and known physical size makes these transformations much simpler, enabling predictive capacities from minimal early visual information. Optical information is inherently ambiguous, and therefore, it is necessary to explain how these prior distributions generate predictions. Here is where the role of prior information comes into play: it could help to interpret and calibrate visual information to yield meaningful predictions of the remaining TTC. The objective of this work is: (1) to describe the primary sources of information available to the observer in parabolic trajectories; (2) unveil how prior information can be used to disambiguate the sources of visual information within a Bayesian encoding-decoding framework; (3) show that such predictions might be robust against complex dynamic environments; and (4) indicate future lines of research to scrutinize the role of prior knowledge calibrating visual information and prediction for action control.

Keywords: 3D perception, calibration, internal models, optic flow, prior knowledge, TTC

OPEN ACCESS

Edited by:
Paul Michael Corballis,
The University of Auckland,
New Zealand

Reviewed by:
Frank T. J. M. Zaal,
University of Groningen, Netherlands
Dennis Shaffer,
Ohio State University at Mansfield,
United States

****Correspondence:***
Joan López-Moliner
j.lopezmoliner@ub.edu

Specialty section:
This article was submitted to Sensory Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *15 December 2020*
Accepted: *28 July 2021*
Published: *23 August 2021*

Citation:
Aguado B and López-Moliner J (2021) Gravity and Known Size Calibrate Visual Information to Time Parabolic Trajectories. Front. Hum. Neurosci. 15:642025. doi: 10.3389/fnhum.2021.642025

INTRODUCTION

Intercepting a ball in a parabolic trajectory before reaching ground level is a fundamental task in different sports: batting a baseball, hitting a high lob in tennis, or heading a football. In those situations, the time at which the interception is possible is very tight, yet our performance is astonishing. Time-to-contact (from now on TTC), that is, the time until an object reaches a location of interest, can provide very useful information that would help anticipate motor programs to solve those tasks.

Study 3

Flexible viewing time when estimating time-to-contact in 3D parabolic trajectories

Borja Aguado
Vision and Control of Action (VISCA) Group, Department of Cognition, Development and Psychology of Education, Institut de Neurociències, Universitat de Barcelona, Barcelona, Catalonia, Spain

Joan López-Moliner
Vision and Control of Action (VISCA) Group, Department of Cognition, Development and Psychology of Education, Institut de Neurociències, Universitat de Barcelona, Barcelona, Catalonia, Spain

Obtaining reliable estimates of the time-to-contact (TTC) in a three-dimensional (3D) parabolic trajectory is still an open issue. A direct analysis of the optic flow cannot make accurate predictions for gravitationally accelerated objects. Alternatively, resorting to prior knowledge of gravity and size can provide accurate estimates of TTC in parabolic head-on trajectories, but its generalization depends on the specific geometry of the trajectory and particular moments. The aim of this work is to explore the preferred viewing windows to estimate TTC and how the available visual information affects these estimations. We designed a task in which participants, wearing an head-mounted display (HMD), had to time the moment a ball in a parabolic path returned at eye level. We used five trajectories for which accurate temporal predictions were available at different points of flight time. Our results show that our observers can predict both the trajectory of the ball and TTC based on the available visual information and previous experience with the task. However, the times at which our observers chose to gather the visual evidence did not match those in which visual information provided accurate TTC. Instead, they looked at the ball at relatively fixed temporal windows depending on the trajectory but not of TTC.

Introduction

The time remaining before an object reaches a point of interest is called time to contact (TTC). To estimate this parameter accurately is of great importance, as it can be used for multiple actions such as avoiding collisions, intercepting moving targets, and, more generally, regulating one's own speed.

Previous literature has focused on the study of TTC for objects under different visual conditions such as objects moving at a constant speed in the frontolateral plane (Bootsma & Oudejans, 1993; Tresilian, 1994) or moving toward the observer (Heuer, 1993; Lee, 1976; Wann, 1996). In contrast, during the past decades, others have focused on the study of gravitationally accelerated objects in free-fall (Lacquaniti & Maioli, 1989; McIntyre et al., 2003; McIntyre, Zago, Berthoz, & Lacquaniti, 2001; Zago et al., 2004), parabolic motion in head-on trajectories (de la Malla & López-Moliner, 2015), or frontoparallel ones (Jörges, Hagenfeld, & López-Moliner, 2018; Jörges & López-Moliner, 2019). However, the estimation of TTC for objects describing parabolas in the more general case has not been systematically addressed. This is probably due to the complex mapping between the distal three-dimensional (3D) trajectory and the projected optic variables. The same optic pattern can be caused by a multitude of sources in the external world (Pizlo, 2001), rendering the interpretation of the real source of stimulation an ill-posed problem known as the inverse problem of vision (Kersten, Mamassian, & Yuille, 2004). This problem would compromise an agent's performance based on optic information alone.

A paradigmatic case in which the computation of TTC for a 3D flying object is key is the so-called *outfielder problem*. In baseball, players known as *outfielders* must catch a flying ball at a specific time and location, avoiding ground contact. The distances involved in this task and the size of the ball render binocular cues and retinal expansion nondiscriminable. Therefore, it is usually assumed that only a reduced set of monocular cues is available to guide action (Cutting & Vishton, 1995; Wilson, Golonka, & Barrett, 2013;

Citation: Aguado, B., & López-Moliner, J. (2021). Flexible viewing time when estimating time-to-contact in 3D parabolic trajectories. *Journal of Vision*, *21*(4):9, 1–16, https://doi.org/10.1167/jov.21.4.9.

https://doi.org/10.1167/jov.21.4.9 Received November 9, 2020; published April 26, 2021 ISSN 1534-7362